Using Random Utility Structural Demand Models to Infer the Value of Quality Information

Celine Bonnet, James Hilger, and Sofia Berto Villas-Boas *

June 30, 2017

Abstract

By conducting a field experiment, we investigate whether consumers value expert opinion labels on wine as a form of reducing asymmetric information about product quality. We use two types of data. First, we use a macro-level monthly-product-store dataset, collected before and after our field experiment, which involves treating a random subset of wine products by displaying expert scores in one store, and comparing sales with wine sales in similar, non-treated stores. Secondly, we use a micro-level panel dataset from the treated store that provides information on products purchased and household characteristics. We combine these data with additional information on products, such as, varietal, region of production, price point relative to other wines, and expert scores. Using the macro-level data for treated and control stores, we estimate a structural random coefficient demand model for wine. With the household micro dataset, we estimate a random coefficient mixed logit demand specification, allowing for consumer heterogeneity in demand for wine. In order to capture the demand for wine, the products are defined as bundles of attributes, including the expert score that is experimentally introduced into the market. We find robust results in terms of consumer valuations for expert scores, using both datasets. In particular, we obtain an implied average willingness to pay (WTP) between 2 (using the macro level dataset for treated and control stores) and 3.2 dollars (using the micro level dataset for the treated store) for an average score of 83. Although not all of the consumer demographics help explain the heterogeneity in the value of expert scores, the wine ratings matter significantly less to consumers that have higher incomes, or are more likely to own a home. Finally, using counterfactual simulations we estimate the changes in consumer surplus resulting from available quality information in the form of expert opinion scores. Using the micro level dataset we find that removing scores leads to significant welfare losses especially for lower income consumers and for men. Overall, using the macro level dataset for treated and control stores, we estimate there to be a significant welfare loss eliminating scores of roughly one percent of total wine revenue in the treated store.

---

*Bonnet: Toulouse School of Economics, INRA, University of Toulouse I Capitole; Hilger: U.S. National Marine Fisheries Service: Villas-Boas: Department of Agricultural and Resource Economics, University of California, Berkeley, CA, 94720. We thank Molly Van Dop for comments; Michelle Scizak, Reed Johnson, and Steve Flaming for help with the data; Will Liu and Jeff Coronado for help with the experimental implementation; and Kyle Birchard, Elizabeta Perova, Grant Chen, Patricia Javier, Katan Patel, and Elizabeth Creed for their excellent research assistance. The authors acknowledge the financial support received from the Giannini Foundation. Corresponding author: sberto@berkeley.edu.

Keywords: Structural random coefficients discrete choice models, Field experiment, Labels, information, expert opinion, wine, product attributes, willingness-to-pay.
JEL Classification: M30, Q18, Q25, Q21, C12, C24.

1 Introduction

This paper uses a randomly assigned introduction of expert opinion scores into the wine market, together with discrete choice modelling specifications (as in McFadden, 1974; Berry et al., 1995; McFadden and Train, 2000; Nevo, 2001; and Train, 2002) to estimate demand and infer the implied revealed preference willingness to pay (WTP) for expert opinion information in the form of scores. We collect the data by designing and implementing a choice experiment in one retail store, where we display expert opinion score labels for a random subset of wines across four weeks. We take advantage of two unique field experimental data and build on the methodological breakthroughs that have arisen in the discrete choice literature when analyzing consumer demand and WTP (see McFadden, 1999 and Train, 2002 for a survey). Using a field experiment, we combine evidence from the revealed preference variation in aggregate and micro level observed choices, to specify and estimate flexible random coefficient mixed logit demand models. This allows us to simulate counterfactual consumer choices with and without expert opinion labels in the wine market. This ultimately estimates the welfare effects of revealing product quality to consumers in the form of expert opinion score information, in a setting of asymmetric information such as the wine market, where consumers know less than producers about the quality distribution of products in their available choice sets.

Given asymmetric information on product quality, consumers must infer quality based on observable attributes at the time of purchase. High quality is typically positively correlated with higher average prices in many markets (Rao, 2005; Shiv et al., 2005). Evidence from blind tastings in the wine market indicates that consumers attribute a positive premium to wines that are perceived as higher quality. Bonnet et al. (2016) shows that uninformed consumers' purchases are consistent with beliefs that high quality is positively related to wine prices. In many markets, experts provide additional insight about the quality of products they evaluate and develop expert ratings or scores that are commonly available to consumers. Producers value expert scores and opinions if they are able to charge higher prices for their high quality products, as they use the higher scores as a product differentiation device to increase market power (e.g., for wine see Ali et al, 2008). On the contrary, whether consumers value expert opinion information in this setting remains an unanswered question.

The first contribution of this paper is to provide macro and micro level structural demand estimates of wine products as a function of wine attributes, such as: price, region of production, varietal, brand, and expert opinion scores, using revealed choices through consumer purchase data. The revealed choice approach has the advantage of 'face validity', as the data are consumers' actual choices when faced with real constraints on their own resources and the products available (Hensher et al., 1998; Whitehead et al., 2008). Consumers consider the internal costs and benefits of their potential choices and experience the consequences of their actions. Choices based on perceived costs and benefits better reflect the values of the population and allow for more valid estimates of willingness to pay. Carson et al. (1996) shows through meta-analysis that estimates from stated and revealed preferences differ. Previous work using the same wine experiment data estimated the reduced form demand effects of wine scores (Hilger et al., 2011). Subsequent work by Bonnet et al. (2016) sought to identify the mechanisms underlying these demand effects, attributing value to different quality grades through the use of a 50 to 100 numeric score. The lack of previous research estimating the value of expert scores, in wine or other markets, is related to the challenge of identifying unbiased demand responses to scores that are uncorrelated with other strategic decisions taken by firms, such as pricing, branding, and product portfolio assortment choices. This challenge is circumvented in this paper, given that the treatment of wines (through revealing their expert scores) is randomly assigned across the potentially scored wine products and is uncorrelated with marketing variables of the wine producers and the retailer. The treated wines had their scores displayed though a label placed on the supermarket shelf underneath the product. When we ran the experiment, we ensured that neither the retailer, nor the producing vineyards, adjusted their marketing variables to take into account the experimentally disclosed score information at the point of purchase.

The second contribution of the paper is to combine the macro and micro level datasets to estimate consumer willingness to pay (WTP) for product attributes and its heterogeneity across observed and unobserved consumer characteristics. We then measure welfare changes due to the introduction of expert opinion information, as a reduction of asymmetric information in the wine market. We develop and estimate a macro (as in Berry et al., 1995 and Nevo, 2001) and a micro level structural model of wine demand specifying a flexible random utility choice framework (as in McFadden, 1974 and 1981, and Train, 2002). That is, we analyze actual response behavior within a designed field experiment for wine retail products to directly estimate the revealed preferences and corresponding WTP measures for wine product attributes. In so doing, we will provide industry participants and policy makers with important information on the efficacy of quality-labels, as well as information on consumer actual wine preferences given consumer characteristics.

Using the experimental variation and detailed product attribute data, we estimate a structural demand model to infer consumer WTP for wine attributes and for expert opinion labels. Given a product-level panel scanner dataset across te treatment and control stores for several months preceeding and during the field experiment, we estimate consumer preferences for products by projecting each wine product into the attribute space, consisting of: price, brand, varietal, and displayed scores. Based on the consumers' choices, we estimate a demand model for the products, and given the demand estimates, are able to assign a dollar value to each product attribute. We also simulate what would have been consumers' choices in the absence of the experimentally displayed scores, estimate the resulting welfare change, and obtain a welfare estimate of revealing information in the form of expert opinion scores, given all other attributes remaining unchanged. In addition, using the micro level dataset for the treated store allows us to formally specify consumer heterogeneity in demand, since we observe consumer characteristics and the resulting consumer choices among the products in the choice set.

We collect the data by designing and implementing a choice experiment in one retail store, where we display expert opinion score labels for a random subset of wines across four weeks. We obtain consumers' demographic characteristics and collect data on their choices among the randomized options presented to each one in the field experiment. Using the panel data, we estimate a discrete choice model for consumer preferences across options given to them in the label field experiment, where a choice is defined as a bundle of attributes: price, expert opinion score, brand, varietal, and region (as in Huber and Train, 2001; Revelt and Train, 1999; McFadden 1974; McFadden and Train 2000; Train, 2002). From the estimated structural demand model parameters from random coefficient mixed logit specifications (Revelt and Train, 1999; Huber and Train, 2001), we obtain estimates for the average WTP for the various specified product attributes. In so doing, this research provides researchers and policy makers with the first estimates of average WTP and its empirical distribution, for attributes among the consumer sample. Additionally, we present novel findings in the heterogeneity of WTP along consumer demographics. Finally, by simulating alternative policy changes in the choice set facing consumers, we obtain estimates of counterfactual individual simulated choices and estimate the resulting welfare changes, measured as changes in the distribution of consumer surplus. We also relate the individual level changes in consumer surplus to the demographic characteristics of the consumers.

Related empirical literature has analyzed the extent to which product quality information affects consumer behavior including branding (Montgomery and Wernerfelt, 1994), mandatory product labeling (Jin and Leslie, 2003; Kiesel and Villas-Boas, 2007), experimental labeling (Kiesel and Villas-Boas, 2008), and advertising (Ackerberg, 2001; Ackerberg, 2003). Closely related to our paper, besides Hilger et al. (2011) that estimates resudec form effects the same experiment of displaying expert scores of wine, are papers by Sorensen and Rasmussen (2004) on the book market and Reinstein and Snyder (2005) on the movie industry. The key identification of the effects of expert opinions on movie demand in Reinstein and Snyder (2005) results from exploiting the timing of movie reviews by Siskel and Ebert. While they find no overall effect of reviews, they show that positive reviews increased box office revenues for narrowly-released movies and dramas, although it remains to be explained why. In the book industry, Sorensen and Rasmussen (2004) find that both positive and negative reviews in the New York Times increase book sales. Our major contribution extending all previous works is that we are the first to assess demand side valuation of expert opinion labels using actual point of purchase decisions of consumers in a field experiment setting. We utilize a flexible discrete choice model (e.g. Berry, Levinsohn and Pakes, 1995; McFadden and Train, 2000; Nevo, 2000; Nevo, 2003; Swait et al, 2004) that incorporates heterogeneity in demand. The framework allows the empirical testing of the null hypothesis that the displayed expert opinion scores are not valued by consumers.

Our macro level estimates suggest that consumers value a score unit by about 2.5 cents, which means that, on average, consumers' WTP is 2 dollars for a bottle of wine with an expert score of the average score of 83. In terms of wine varietals, consumers are willing to pay one dollar more for California wines over other wines in the sample. Other varietals are also positively valued, with the Chardonnay varietal having the highest WTP of 2.9 dollars a bottle, and the lowest estimated WTP for the Merlot varietal. Using counterfactual simulations, we estimate that eliminating expert scores leads to a significant welfare loss of 179 dollars, representing 1.1 percent of total revenue from wine sales in this store.

From the micro level household dataset, we find that consumers are willing to pay an average of 3.25 dollars more for the quality information from the score label for a wine with an average score of 83. From the consumer demographic characteristics we find that there is heterogeneity in the WTP along consumers' likelihood of owning a home and their income. Using counterfactual simulations of removing expert opinion scores from the choice set attributes, we estimate changes in choices that imply significant consumer surplus losses, especially for consumers with lower incomes and those who do not own a home.

The remainder of the paper is structured as follows: Section 2 presents the data and experimental variation used. Section 3 specifies the structural demand model and derives how to obtain the implied estimates of consumer valuation for expert scores using aggregate choice data for treated and control stores and the household level micro dataset with observed consumer characteristics. Section 4 presents and discusses the structural estimates using both macro and micro level data, and presents the estimates of WTP using both data sets. In Section 5, we lay out the methodology and present the results of counterfactual simulations, to estimate average and heterogeneous consumer surplus changes due to displaying expert scores. Finally, Section 6 contains closing remarks.

2 The Experiment and the Data

In April 2006, wine ratings from a proprietary wine scoring system were displayed in the treatment store for four weeks. We labeled 101 wine products with scores in the treated store, which corresponds to displaying scores for about 14% of the wines in the consumers' choice set. Each label features the name of the proprietary scoring system and the wine's score that. In theory, wine scores range from a low value of 50 to the high score of 100. However, scores less than 70 are not released by the rating agency.

Figure 1 displays the kernel density of the score distribution for treated wines in the treated store and the kernel density of the score distribution of the unlabeled products sold in the control stores, given that we can see the same products in the control stores. Given a Kolmogorov-Smirnov (KS) test for equality of the distributions, we cannot reject that the distribution of scores are equal across the treated and control stores. Therefore, there exists a nice match in the distribution of labeled wines across our sample.

The treated store is in the same marketing division as a set of 38 potential control stores. The pricing, promotions, and display layouts are common among all of the stores in a marketing division, leading to a good balance of observable determinants of quantities of wine sold, originating from the retail marketing strategy.

2.1 The Store Level (Macro) Data

We use a weekly scanner dataset for treated wine products in the treated and in the four control Northern Californian stores (among the 38) that match the treated store in terms of pre period trends in labeled wines. The data provide a unique wine product code identifier (UPC), the name of the wine (including varietal), the number of bottles sold, the pre-discount price paid, and any retail discount pricing offered. We aggregate the weekly sales data to the month-level for each store to generate the total number of bottles sold per month, the average shelf price, the average price paid (the shelf price net of discounts), and whether a bottle of wine was discounted during the given month. Pricing and discounting for each product are common for all of the stores in the data. Moreover, we made sure that wine pricing was not updated due to the selection of products into labeled and unlabeled status, and prices were not differentially updated in the treated store due to our experiment.

For those wines for which proprietary wine score data exist, we merge the wine score data into the scanner data. In addition, we collect a detailed product attribute dataset, identifying the brand of the wine product, varietal, type (red, white, or other), regional designation, and imported status, which we merge with the scanner dataset.

Summary statistics for the aggregate choice macro data set used in the analysis are reported in Table 1. We report descriptive statistics for the treated store and for the control stores in the first and second columns, respectively. The summary statistics report average quantity sold during the pre-treatment and during the treatment month (April), along with the standard deviations. In the treated store, 17 bottles of the labeled wines were sold on average, while 10 bottles were sold in the control stores during the month of March. Average prices in March and April are 11 dollars for treated wines. The averages are not statistically different between treated and control stores. Approximately 90 percent of the wine consumers purchase is discounted in March and in April, across both treated and control stores.

The bottom part of Table 1 reports the average and the standard deviation of scores for treated and control stores. Average scores are around 83.13 for treated wines. Additionally, Figure 1 attested that the treated and control stores had similar average scores along with very similar score distributions. In both the treated store and control stores, 58 percent of the treated group are classified as red wines. The proportions of white wine are also similar across the treated and controls stores. In the treated store, we have 2562 observations for monthly sales of treated wine products. In the four control stores, the number of observations total to 11058.

Given the total quantity Q of wine sold monthly by store, we construct product market shares by dividing each product's quantity sold by the total quantity Q. At most, a wine product represents 8 percent of total monthly wine sales in a store, and the density of market shares are very similar between treated and control stores. To estimate the causal effect of revealing expert scores on consumer demand and valuation, it is crucial that there are similar pre-period trends across treated and control stores for products in the analysis, with respect to quantity sold and market shares. Figure 2 shows that trends in the sum of the monthly market shares of labeled products are quite similar in the treated and control store. The similarity in trends allows us to investigate the causal effects of the display of labels on treated wines on demand choices and infer from that WTP for those displayed scores.

2.2 The (Micro) Household Level Panel Data for the Treated Store

Summary statistics of the household micro level data set are presented in Table 5, which is organized in two panels. In Panel A, the demographic makeup of the panel is compared to the total California population. In Panel B, we present average choices in the panel group with and without demographics for the labeled and unlabeled wines, before and after the treatment weeks. The household panel follows 4,754 wine purchasing households, 3,590 of which we have demographic information. The purchase dataset features 31,361 observations of wine product weekly purchases over a two year period for a total of 24,610 (product-household-week) observations for households with demographic information and 6,751 observations for households who did not report demographic information.

Panel A of Table 5 presents the proportion of males and females, empirical distributions for age, education, and income, and racial composition of the white (including Hispanic) and non white population for California (column 1), and for the subset of the purchase panel with demographic information (column 2). The consumer sample gender statistics are similar to the CA population. However, in our panel dataset, ages "17 or younger" are not represented, as they cannot buy any alcohol in stores, and the group is therefore underrepresented compared to the California population. The "60 or older" age group is overrepresented in the sample, suggesting that purchase data are skewed towards older populations. Income levels in the purchase panel are also higher in general than the California populace. Finally, almost all consumers in the purchase panel are likely to own a house and less likely to rent. We have no education nor racial composition data for the sample.

In Panel B of Table 5, we present purchase panel summary statistics of choices choices made by the consumers with demographic information (column 1) and for the full sample in column 2. The first four rows report the average quantity of labeled and unlabeled products purchased by the household in the pre- and post-treatment weeks. In the sample with full demographic data, we see that the average purchases of labeled wines increase from 1.3 to 1.5 bottles and average number of unlabeled wine purchases decrease from 1.47 to 1.42 bottles. For all households in the panel, labeled wine purchases increase from 1.329 to 1.446 bottles and unlabeled wine purchases drop from 1.449 to 1.438 bottles.

Panel B also reports the average share of purchased labeled options pre- and post-treatment. The share variable is constructed as follows: every week, each household can potentially buy more than one wine product among the choice set. In the data, it is not always the case that a household only buys one of the wine products. To estimate demand, we construct for each household the share by week of each of the products purchased as the ratio of the quantity of product j purchased that week, divided by the total potential products a household could purchase a week among the choice set (analogous to the potential market using aggregate data in discrete choice in Berry, 1994 and Berry et al., 1995), where the potential number of products that could be purchased is defined as the maximum number of wine products that a consumer ever purchased in the two years of data. The sum of the share of products purchased is equal to the "inside share" whereas the remainder is the "outside share". The outside option consists of not buying the sample products that week. Going back to the table, in the rows "Share Labeled Pre" we see that, on average, the "inside share" is 16.8 percent for both columns and increases post-treatment for the demographic sample, meaning that the household average probability of not buying in a week is about 83 percent.

Finally, in the bottom rows of Panel B, we report that the average price of the chosen options pre- and post-treatment is roughly 9 dollars, for both the full and the demographic data sample. A high percentage of the chosen products are purchased at discount (above 90% for both groups of households), 59 % of chosen products are white wines, 37 % red, and the percentage of California wines purchased in the pre period weeks is 73.5 %.

3 Consumer Demand Model

Using a store macro level and a household micro level panel dataset, along with product characteristics and consumer characteristics data, we estimate flexible specifications of discrete choice structural revealed preference models of consumer demand. First, we derive the macro level model approach and then the micro level model approach for estimating consumers' willingness to pay (WTP) for product labels with expert opinions on product quality. Modeling consumer choice as the demand for product bundle of observable attributes, we are able to estimate a dollar value for each attribute. Values of consumers' WTP for expert information are empirically estimated through the addition of an expert opinion attribute to the product space, which is introduced through the field experiment.

3.1 Structural Demand Model using the Macro Level Data

Taking advantage of these unique aggregate choice data within the field experiment, we identify consumers' valuation of the expert opinion label with a discrete choice model approach (McFadden, 1974; Train, 2002). We define the consumer product as a bundle of perceived product attributes, which allows us to compute consumer's willingness to pay for additional labeling information in a straightforward way. In this context, we further define product-specific information provided through expert opinion labels as additional or differentiated product attributes. The discrete choice model (e.g. Berry, Levinsohn and Pakes, 1995; McFadden and Train, 2000; Nevo, 2000; Nevo, 2003; Swait et al, 2004) also offers flexibility in incorporating heterogeneity.

Starting from a random utility framework (e.g. McFadden, 1974; Train, 2002), where both the product attributes as well as a random term are assumed to enter linearly, the utility from consuming a certain product j can be described as:

$$U_{ijt} = a_j + a_t - \alpha_i p_{jt} + X_{jt}\beta + \gamma T_{jt} Score_{jt} + \xi_{jt} + \varepsilon_{ijt}, \tag{1}$$

where a product is defined as a particular wine UPC sold at a certain store, a_j is a product (UPC-Store) fixed effect capturing the intrinsic preference for product j and a_t is a time fixed effect. The shelf price of product j at time t is denoted by p_{jt} and the marginal utility of price is α_i. In γ_i, we measure consumers' average marginal utility for the labeled score experimentally displayed on product j. T_{jt} is a dummy variable that is equal to the one during the treatment period in the treatment stores and equal to zero otherwise, and $Score_{jt}$ is the value of the displayed score for product j. A treated store indicator and treated weeks indicators are included in X_{jt}. The term ξ_{jt} accounts for monthly changes in factors such as shelf space or positioning of the product among others that affect consumer utility, and are observed by consumers and firms but not by the researcher. Lastly, ε_{ijt} is an i.i.d. type I extreme value distributed error term, that capture consumer idiosyncratic preferences.

To allow for category expansion or contraction, we include an outside good (no-purchase option), indexed by $j = 0$, whose mean utility is normalized to zero. Therefore, its utility is given by the idiosyncratic term only:

$$U_{i0t} = \varepsilon_{i0t}. \tag{2}$$

Let the α_i coefficient vary according to

$$\alpha_i = \alpha + \sigma v_i,\ v_i \sim N(0, 1), \tag{3}$$

where α and σ are parameters to be estimated.

As in Nevo (2000), we rewrite the utility of consumer i for product j as:

$$U_{ijt} = \delta_{jt}(p_{jt}, X_{jt}, \xi_{jt}; a, \alpha, \beta, \gamma) + \mu_{ijt}(p_{jt}, v_i; \sigma) + \varepsilon_{ijt}, \tag{4}$$

where δ_{jt} is the mean utility and μ_{ijt} is the deviation from the mean utility that allows for consumer heterogeneity in the marginal utility response to price, product attributes, and the treatment.

Let the distribution of μ_{ijt} across consumers be denoted by $F(\mu)$. Then the aggregate probability S_{jt} of product j at month t across all consumers is obtained by integrating the consumer level probabilities:

$$S_{jt} = \int \frac{\exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{n=1}^{N} \exp(\delta_{nt} + \mu_{int})} dF(\mu). \tag{5}$$

When estimating demand, the goal is to derive parameter estimates that produce product market shares close to the observed shares. This procedure is non-linear in the demand parameters, and prices in general enter as endogenous variables, although prices are not set at the product-month-store level. Instead, prices are set at the wine-price-marketing division level, which covers all of the stores in the sample. We will treat prices as exogenous determinants of demand given that prices are decided at a more macro level than by an individual store.

We estimate the random parameters logit demand model from product (UPC-store) monthly market shares using the GMM-estimator proposed by Berry, Levinsohn, and Pakes (1995) and Nevo (2001). We allow for consumer heterogeneity in the valuation of product characteristics, which can be modeled with a normal random variable, given that we do not have demographic variables.

We follow Berry (1994), who constructs a demand side equation that is linear in the parameters to be estimated. This follows from equating the estimated product market shares[1] to the observed shares and solving for the mean utility across all consumers, defined as:

$$\delta_{jt}(a, \alpha, \beta, \gamma) = a_j + a_t - \alpha p_{jt} + X_{jt}\beta + \gamma T_{jt} Score_{jt} + \xi_{jt}. \tag{6}$$

For the random coefficient logit model, solving for the mean utility (as in Berry 1994) has to be done numerically (see Berry, Levinsohn, and Pakes, 1995; and Nevo, 2001). Once this inversion has been made, one obtains equation (6) which is linear in the parameter

[1]For the random coefficient model the product market share in equation (4) is approximated by the Logit smoothed accept-reject simulator.

associated with all wine attributes. The estimates are obtained by a fixed effects OLS regression. We let $\alpha, \beta, \gamma, \sigma$ be the demand side parameters to be estimated, where the linear parameters are $(a, \alpha, \beta, \gamma)$ and σ is the price non-linear random coefficient parameter. In the random coefficient logit model, the parameters are obtained by feasible Simulated Method of Moments (SMOM) following Nevo's (2000) estimation algorithm, which requires equation (6).[2]

The demand model represents consumer choice between different wine products over time, where a product is perceived as a bundle of attributes, including expert scores and price. A product-store fixed effect is included to capture constant observed and unobserved product (UPC-Store) factors that affect demand. The econometric error that remains in ξ_{jt} will therefore only include the (non-product specific) changes in unobserved product characteristics such as unobserved consumer level determinants of demand.

3.1.1 Estimating Average and Heterogeneous Marginal Utility and WTP

Using the dataset of product choices, we estimate a random coefficient Logit choice model given by equation (6). We obtain estimates of average WTP for the labels by dividing the estimates for γ by the average marginal utility of price α, given store-level observed purchases.

Given the demand estimates, we obtain estimates for the predicted market shares of each

[2]The aim is to concentrate the simulated GMM objective function such that it will be only a function of the non-linear parameters. By expressing the optimal vector of linear parameters as a function of the non-linear parameters and then substituting back into the objective function, it can be optimized with respect to the non-linear parameters alone.

wine product j for draw i by:

$$\hat{s}_{jit} = \frac{\exp(\hat{a}_j - \hat{\alpha}_i p_{jt} + X_{jt}\hat{\beta} + \hat{\gamma} T_{jt} Score_{jt})}{1 + \sum_{n=1}^{N} \exp(\hat{a}_n - \hat{\alpha}_i p_{nt} + X_{nt}\hat{\beta} + \hat{\gamma} T_{nt} Score_{nt})}. \tag{7}$$

The market share is obtained by averaging (7) over all the normal draws, that is, given by $\hat{s}_{jt} = \frac{1}{R}\sum_{i=1}^{R} \hat{s}_{ijt}$, where R are the number of normal draws of v_i.

3.2 Micro Data Set based Structural Demand Model

Using the individual consumer panel data set with consumer-specific choice information and consumer demographics enables us to consider and estimate a specification of heterogeneous preferences for product attributes directly into the discrete choice model. Given that each household can choose more than one wine product among the available choice set, we estimate the probability of a series of choices in each choice occasion for each household. Starting from a random utility framework (as in McFadden, 1974; McFadden and Train, 2000; and Train, 2002) where the product price p, the product attributes X, and the random term ε are assumed to enter linearly, the utility from consuming a certain product can be described as in equation (3.1). For the scores and attributes of product j, the marginal utility that individual i places on these attributes are specified by (β_i, γ_i), respectively, as:

$$\theta_i = \theta_0 + \theta_1 D_i, \tag{8}$$

which indicates that the coefficients vary according to the consumer's observed demographics D_i, and allows for the fact that different decision makers may have different preferences.

We allow marginal utility for price to vary according to:

$$\alpha_i = \alpha + \sigma v_i, \tag{9}$$

where v_i is a normal random variable capturing any heterogeneity. If ε_{ji} are assumed to be independently, identically extreme value distributed (type I extreme value distribution), the choice probabilities constitute a random coefficient Logit model if α_i is specified as in (9), which offers flexibility in incorporating consumer heterogeneity with regard to wine price.

This choice model offers flexibility in incorporating consumer heterogeneity over product attributes as a function of D_i, while also allowing for random determinants of heterogeneity in marginal utility of price, via v_i. This modeling approach, combined with this unique field experiment and resulting data variation for product choices, allows us to estimate consumers' average valuation for product attributes (as in Revelt and Train, 1999; Hubert and Train, 2001) along with the complete distribution of valuations by the consumers in the purchase dataset.

The probability that good j is chosen is the probability that good j maximizes the consumer i's utility and results in the predicted probabilities equal to:

$$Prob_{ji} = \frac{e^{(X_j, TScore)\theta_i + \alpha_i \mathrm{p}_j}}{\sum_{k=0}^{N} e^{(X_k, TScore)\theta_i + \alpha_i \mathrm{p}_k}} \tag{10}$$

where α_i is the marginal utility with respect to price, and $\theta_i = (\beta_i, \gamma_i)$ contain the marginal utilities relative to the remaining attributes X and the score treatment $TScore$ for consumer i. The mean utility of the no purchase option is normalized to zero. The attributes and price variables for that alternative are set equal to zero, which implies that equation (10) becomes:

$$Prob_{ji} = \frac{e^{(X_j, TScore)\theta_i + \alpha_i \mathrm{p}_j}}{1 + \sum_{k=1}^{N} e^{(X_k, TScore)\theta_i + \alpha_i \mathrm{p}_k}}. \tag{11}$$

In the data, we see the same household making several choices during the same week. Therefore, we cannot consider micro-level binary purchase decisions. Instead, we consider the choices a household makes during a certain week as being "shares" of the purchases allocated to one of the products in the choice set, including no purchase, as we did in the model using the store level data. In order to estimate the above equation, we follow Berry (1994). For each choice occasion of individual i, we normalize the probability of choosing one particular product $Prob_{ji}$ by the probability of purchasing none of the wine alternatives presented at the store that week. The empirical analog of $Prob_{ji}$ in week t, given by (11), is the share of product j purchased in week t by consumer i. Defining the maximum number of total items a household could purchase M_{jt}, calibrated as the maximum one household in the sample ever purchased of wine in each week of data, we obtain the market share of product j purchased by household i in week t as q_{jit}/M_{jt}. This choice model considers that each household chooses to purchase different proportions/shares of the available options (or purchase nothing) each week.

Following Berry (1994) by the log of (11) minus log of (s_{0i}) for consumer i, we obtain an equation to estimate that is linear in the average marginal utilities and mixing parameters and is non linear in the random coefficients as:

(12)

$$ln(Prob_{ji}) - ln(S_{0i}) = ln(\frac{e^{(X_j,TScore)\theta_i+\alpha_i \text{Price}_j}}{1+\sum_{k=1}^{N} e^{(X_k,TScore)\theta_i+\alpha_i \text{p}_k}}) - ln(\frac{1}{1+\sum_{k=1}^{N} e^{(X_k,TScore)\theta_i+\alpha_i \text{p}_k}})$$

(13) $$ln(Prob_{ji}) - ln(S_{0i}) = X_j\beta_i + \gamma_i T_j Score_j + \alpha_i \text{p}_j,$$

where α_i is defined in equation (9) and θ_i is defined in (8). In the heteregeneous conditional logit model (12) can be estimated with a panel regression, using observations for all weeks and product shares of all households. The random coefficient mixed model equation (12) is estimated following Berry et al. (1995) and Nevo (2001) using Simulated Method of Moments (SMOM) following Nevo's (2000) estimation algorithm.

3.2.1 Estimating Average and Heterogeneous Marginal Utility and WTP

Using the dataset of product choices, we estimate a random coefficient mixed logit choice model of consumer demand for the products in the sample, given by (11). Each product is defined as a bundle of attributes, and we therefore estimate the parameters α and the (β, γ). Not only we can estimate average marginal utility for a certain attribute $x, TScore$, but we can also estimate heterogeneity θ_i in the marginal utility for each consumer i in the sample by adding D_i as mixing parameters into the heterogeneity specification directly, as given by equation (8).

Given that the expected value of θ, conditional on a given response Y_i of individual i and a set of alternatives characterized by X_i at occasion t, is given by:

$$(14) \qquad E[\theta|Y_i, Xi] = \frac{\int \theta \prod_{t=1}^{T_i} \prod_{j=0}^{J} \left[\frac{e^{(X_{ijt}, T_j Score_j)\theta_i + \alpha_i \text{Price}_{jt}}}{1+\sum_{k=1}^{N} e^{(X_{ikt}, T_k Score_k)\theta_i + \alpha_i \text{Price}_{kt}}} \right]^{Y_{ijt}} f(\theta|\theta_0, \theta_1, \alpha, \sigma) d\theta}{\int \prod_{t=1}^{T_i} \prod_{j=0}^{J} \left[\frac{e^{(X_{ijt}, T_j Score_j)\theta_i + \alpha_i \text{Price}_{jt}}}{1+\sum_{k=1}^{N} e^{(X_{ikt}, T_k Score_k)\theta_i + \alpha_i \text{Price}_{kt}}} \right]^{Y_{ijt}} f(\theta|\theta_0, \theta_1, \alpha, \sigma) d\theta},$$

then (14) can be thought of as the conditional average of the coefficient for the sub-group of individuals who face the same alternatives and make the same choices (Train, 2002). For each individual i, we follow Revelt and Train to estimate a certain attribute's β_i by simulation according to the following:

$$(15) \qquad \hat{\theta}_i = \frac{\frac{1}{R} \sum_{r=1}^{R} \theta_i^{[r]} \prod_{t=1}^{T_i} \prod_{n=1}^{N_{ti}} \prod_{j=0}^{J} \left[\frac{e^{(X_{ijt}, T_j Score_j)\theta_i + \alpha_i^{[r]} \text{Price}_{jt}}}{1+\sum_{k=1}^{N} e^{(X_{ikt}, T_k Score_k)\theta_i + \alpha_i^{[r]} \text{Price}_{kt}}} \right]^{Y_{ijt}}}{\frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T_i} \prod_{j=0}^{J} \left[\frac{e^{(X_{ijt}, T_j Score_j)\theta_i + \alpha_i^{[r]} \text{Price}_{jt}}}{1+\sum_{k=1}^{N} e^{(X_{ikt}, T_k Score_k)\theta_i + \alpha_i^{[r]} \text{Price}_{kt}}} \right]^{Y_{ijt}}}$$

where $\alpha_i^{[r]}$ is the r-th draw for individual i from the estimated i's distribution of α. The resulting estimates of each consumer's willingness to pay for a particular attribute, including the experimentally introduced expert score, are obtained as the ratio of θ_i and the average marginal utility with respect to price α. We can therefore recover not just the average WTP but also the distribution of the WTP in the sample of consumers, and can find standard errors through an application of the Delta Method. Finally, we relate the estimated willingness to pay (WTP_i) to each consumers' demographics by estimating the equation:

$$(16) \qquad WTP_i = \delta_0 + \delta_1 D_i + \epsilon_i$$

where WTP_i is a vector of all the consumers' individually estimated willingness to pay for the attribute of interest, D_i are the demographic characteristics of consumer i, and δ_0, δ_1 are parameters to be estimated.

4 Results

Within this section, we first present estimates using the macro-level dataset followed by estimates with the micro-level dataset. For both, we discuss flexible random coefficients logit specification results in terms of marginal utilities for all the wine products. Finally, given demand, we estimate the average WTP for observable wine attributes using the macro level data. Using the micro level purchase panel results, and taking advantage of the household level attributes, we finish by presenting consumer heterogeneity in the WTP estimates.

4.1 Structural Demand Results - Random Coefficient Logit Marginal Utilities using Macro Data

The demand model estimates obtained by GMM, as in Reynaert and Verboven (2014), Nevo (2000) and Berry et al., (1995), are presented in Table 3. In column (1), we estimate demand with respect to price, a California region dummy, a discount dummy, treatment store and period interactions, and the coefficient of interest, associated with the "Score X Treated Store X Treated Period" variable in the first row. In column (2) we estimate demand as in column (1), with the inclusion of varietal fixed effects. This specification is motivated by the significant marginal utilities obtained in the logit specification. We obtain the desired coefficients to estimate the WTP in the next section by dividing the marginal utilities of all the attributes by the marginal utility of price, which is estimated to be -0.199 and is significant. The marginal utility of price also has significant heterogeneity, given that the estimated sigma is significantly different from zero, found in Table 3 as “SD Price”, with a value of 0.1. Ihe marginal utilities of a wine from California is positive, as is the marginal utility of most varietals, given the positive and significant estimates in column (2) of Table 3. Once again, consumers place a positive marginal utility on expert scores displayed during the treatment period at the treated store, given the significant point estimate of 0.005.

4.2 Results in terms of Heterogeneity Using Household Panel Dataset

We present the results from the choice estimates originating from a conditional logit specification as a first step in understanding whether the mean and deviation from the mean of the stated marginal utilities for the product attributes are significant, as a function of consumer demographics. Then, we explore a more flexible random coefficient choice model, allowing for the heterogeneity to vary from the average marginal utility in a random fashion. Finally, we include demographics D_i as mixing parameters directly and estimate the random coefficients mixed logit model. Given the choice estimates, we recover the implied marginal utilities for the attribute of interest of this paper: the expert opinion score.

Each consumer’s θ_i is computed as a conditional average of θs of consumers similar to them, in that they make similar sequences of choices when presented with the same options in the experimental design, and that they have similar D_i. Each consumer’s WTP for an attribute of interest, such as expert scores, is then obtained as the ratio between the θ_i and the marginal utility of price α. The variation in estimated individual departures from the average WTP can be either purely random, or they can be due to the fact that consumers have similar characteristics. This is investigated by correlating the estimated WTP_i with consumers’ demographics.

4.2.1 Heterogeneous Micro Data Based Estimates

In Table 6, we present the estimates of the choice model specification where θ_i are given by (8). The dependent variable in all of the columns is the log of the odds ratio relative to the outside option, namely, $ln(s_{ji}) - ln(s_{0i})$. In all the specifications, we include brand and week fixed effects, controlling for brand constant characteristics that may affect average choice behavior and controlling for anything that changes weekly that is common to all consumers or options.

In column (1) of Table 6, the right hand side variables are the price, the variable "Treatment" (which equals one if the observation deals with a treated wine and a treated week, and is equal to zero otherwise), and interactions "Treatment X Score" and "Treatment X Score X D". D includes the demographic information of income, age, and gender (Income, Age, Female). In column (2), we add additional demographic information to D, where we interact "Likelihood of Owning Home" with "Treatment X Score".

From the estimates in column (1) and (2), we see that the coefficient of price is negative and significant, meaning that a high price lowers the marginal utility of purchasing wine products. The marginal utility of the white wine type is positive and significant, whereas for the other attributes the point estimates are not significant. In general, marginal utility due to the "Treatment" is negative but not significant, and on average the higher the score, the larger the marginal utility, given the point estimate of "Treatment X Score" being positive. In column (2), we see that on average people value expert scores, given the positive and significant point estimate for "Treatment X Score", with a value of 0.046. However, the disclosure of scores is not valued by those likely to own a home, given the negative and significant coefficient associated with "Treatment X Score X Likelihood of Owning Home" of -0.028. This implies that, for those not likely to own a home, the expert opinion information and high-scoring wines have significant value. None of the other demographic interactions are significant.

While this demand choice model specification includes directly observed consumer characteristics when estimating consumer taste parameters, we next turn to a more flexible choice specification, where we allow the average taste parameters to vary randomly for the consumers in a mixed logit specification, and not just as a function of a set of observable consumers' characteristics.

4.2.2 Random Coefficient Mixed Logit Choice Estimates

In the first column of Table 7, we present the estimates of the random coefficient mixed Logit choice model, where the price coefficient is allowed to vary as a random coefficient in columns (1) and (2). In column (2) we allow the consumer demographics to interact with the variable "Treatment X Score". The dependent variable in all of the columns is the log of the odds ratio relative to the outside option, namely, $ln(s_{ji}) - ln(s_{0i})$. In all columns, the right-hand-side variables are the price, the indicator variable for treated wines during the treated weeks "Treatment", and interactions "Treatment X Score" and "Treatment X Score X D", where D includes the demographic information "Income", "Age", "Female", and "Likelihood of Owning Home".

The price coefficient is negative and significant in all columns of Table 7 and has a similar magnitude as the marginal utility estimates of price for the conditional logit specifications in Table 6. Therefore, from the estimates, we see that a high price lowers the marginal utility of purchasing an alternative. The average marginal utility of the white wine attribute is positive and significant. Finally, being labeled with a larger score implies a higher marginal utility, given the positive coefficients of the interaction "Treatment X Score" and it is significant in the random coefficient mixed logit specification in column (2), similar to the findings in Table 6.

There is no significant heterogeneity in the marginal utility of price, given that the estimates for the coefficients of the standard errors of price in column (1) and (2) are not significant.[3] However, in column (2) there is significant heterogeneity of the marginal utility of labeled scores relative to the likelihood of owning a home.

As a conclusion, there is significant heterogeneity in the marginal utility of the expert scores interacted with the likelihood of owning a home, given the negative and significant coefficient from the estimation, which is consistent with the findings in Table 6. Moving forward, we allow heterogeneity based on demographics and not based on random factors, given that those were not significant. We therefore use the conditional logit estimates given in column (2) of Table 6 to then recover the implied WTP and perform policy simulations.

4.3 WTP for Wine Attributes and Expert Opinion Experimentally Displayed Scores

4.3.1 Macro Data Set Evidence

We divide all of the marginal utilities by the average marginal utility of price that, following Train (2003), is assumed as a fixed parameter, to obtain estimates of average WTP as

[3] In additional specifications, we also found that there is no significant unobserved random heterogeneity in the marginal utility of scores for labeled wines, given the positive but not significant coefficient for the standard errors of the "Treatment X Score" marginal utility. These are not reported in Table 7 to save space.

reported in Table 4. Based on these estimates we find an overall positive average consumer willingness to pay for expert opinion scores, with a WTP 2.5 cents for each score point. Given that the average score displayed is 83, this means that consumers are willing to pay about 2 dollars more for the average scored wine, than had they not received the posted score information. Other attributes amount to significant WTP estimates. In particular, consumers are willing to pay nearly a one dollar premium for California wines. With respect to varietals, the Chardonnay has the highest WTP of 2.9 dollars, followed by Cabernet with a WTP of 1.8 dollars, and a WTP of 1.1 dollars for both Pinot and the Gris/Gewurztminer combined type. Merlot has a positive, albeit not significant, WTP estimate based on the purchases of consumers in our data.

Assessing the WTP of expert opinions in the wine market is a previously unexplored avenue in the literature. However, we could find some WTP estimates for some other characteristics of wine, deduced from consumer questionnaires. Louriero (2003) used a survey on 100 wine consumers and found a small WTP for Colorado wines and environmentally friendly wines, 4 and 17 cents respectively. Bazoche et al. (2008) found through an experimental study of French wines that the WTP for the effect of information about the pesticide use in farming is 2.3 on average (roughly 3 dollars). Sellers-Rubio and Nicolau-Gonzalbez (2016) show that the WTP for sustainable Spanish wines is +12.9% of the price of non-sustainable wines, corresponding to 1.4 dollars for an 11 dollar bottle of wine. In the Hong Kong wine market, Song et al. (2015) used an hedonic price function to reveal a positive WTP for red wines with respect to white wine, for Old World wines with respect to New World ones, and for well-known grape varieties.

4.3.2 Micro Based Average and Heterogenous Estimates of WTP

Given the estimated model parameters in column (2) of Table 6, we estimate the average and the distribution of the consumers' individual marginal utilities and resulting WTP_i with respect to expert scores. The point estimate of the average WTP is about $0.046/0.014 = 3.26$ dollars for the labeled scores attribute. Given that this attribute is associated with an average score of 83.6, then the average WTP per score point is $\frac{3.26}{83.6}$, which is equal to 4 cents per quality score point.

Given (8), θ_i is estimated for each individual, and then divided by the marginal utility of price α to obtain each WTP_i. Heterogeneity in the WTP for expert scores is formally investigated by estimating equation (16), a linear regression of the estimated individual WTP and all the demographic characteristics of the consumers. The estimates are reported in Table 8. While the likelihood of owning a home, age, and gender are not significantly correlated with the WTP for labeled scores, a consumers' income is negatively and significantly correlated with WTP.

5 Choice Changes and Welfare Changes in Counterfactual Policy Simulations

Finally, we ask the counterfactual question of what would happen to consumers' choices and to consumer welfare, *ceteris paribus*, were there to be no score information revealed to consumers. To answer this question, we perform simulations and compute the maximizing utility choices for each consumer in this counterfactual scenario. In so doing, we are able to simulate consumers' new choices and estimate the distribution of changes in consumer surplus. To assess who loses and who wins, we project the changes in consumer surplus on consumers' demographics in the final step. This section is organized as follows: we perform such simulations using the macro level and then the household level micro dataset.

5.1 Estimating Macro Level Consumer Welfare Changes in Policy Simulations

Estimates of changes in consumer surplus (CS) are derived through simulation of consumer choices under a counterfactual attribute composition of their choice sets. These correspond to a respondent's compensating variation for a change in product attributes (Small and Rosen, 1981). The expected consumer surplus, CS_i, is defined as

$$CS_i = \frac{1}{|\alpha|} ln \sum_j e^{a_j + a_t + X_{jt}\beta - \alpha_i price_{jt}}, \tag{17}$$

where α denotes the average marginal utility of price. We estimate the consumer surplus for the baseline choices when scores are displayed and the consumer surplus for the next best alternatives consumers choose when there is no longer score information available. The distribution of estimated changes in consumer surplus are then obtained, and we estimate the average change in consumer surplus from eliminating the expert scores in the treated store as:

$$\Delta CS = \sum_i \Delta CS_i. \tag{18}$$

where CS_i is given by (19) and ΔCS is the Total Consumer Surplus without Scores less the Total Consumer Surplus with Scores.

As we estimated, the average scored wine WTP amounted to 2 dollars, testifying to the value consumers place on the expert opinion scores displayed. Here we perform a welfare analysis, resulting from having introduced scores into this market using the flexible random coefficients demand model. The procedure is to estimate consumer surplus for choices made when scores are available, to simulate what consumers' choices would be in the absence of scores, and then to estimate the resulting counterfactual consumer surplus. The difference in surplus amounts to the welfare change due to eliminating scores. In other words, a negative change in surplus when eliminating scores means that scores are significantly valued by consumers, and by adding the total change in consumer surplus, we obtain the total value of displaying the scores in this particular market.

First, we estimate the product level market shares of the choices, taking as given the estimated parameters for the baseline scores. Then we predict the choices made when no scores are available. We see that removing scores had the biggest effect on increasing the outside option. Finally, we estimate the changes in consumer surplus by comparing the baseline and the counterfactual scenarios compensated variation given the baseline and the simulated aggregate choices.

Figure 4 displays the estimated kernel density of consumer surplus for the baseline choices, as well as the estimated kernel density of consumer surplus when consumers are faced with the same wine options without expert scores, observing only the price and product constant attributes (such as the brand and varietal). We see a shift to the left of the density without scores, meaning that there is a higher mass of consumers with lower surplus in the simulated counterfactual than at the baseline.

Overall, the visual evidence suggests that this policy experiment of removing expert scores has a net welfare loss. This indicates that the expert scores provided an increase in consumer surplus and welfare. Moreover, we estimate there to be significant consumer surplus losses of 178 dollars, representing 1.1 percent of the revenues in this market.

5.2 Simulating Micro Level Respondents Counterfactual Choices

We repeat the above approach using the micro level dataset. For the counterfactual scenario, we keep consumer preferences unchanged. In practice, this means that the marginal utility parameters do not change from the baseline model prior to the simulations. Given the data on the attributes pre-simulation from (11), we estimate the probabilities of each alternative being chosen in each case by all consumers, and obtain the predicted pre-simulation baseline choices for all consumers. Then, we change the vector of attributes under the counterfactual scenario considered, defined as $\tilde{X}$, and recompute the probabilities of the choices that each consumer would make under this scenario for all cases, using the new attributes. For example, simulating no score labels means that all products are indistinguishable along the score level differentiation attribute in this counterfactual scenario. In practice, this means that $X_{ij,score} = 0, \forall i, j$ products and consumers, which also implies that all interactions with that attribute are zero in the scenario.

5.2.1 Estimating Consumer Welfare Changes in Policy Simulations

Estimates of changes in consumer surplus (CS) are derived through the simulation of consumer choices under counterfactual attribute composition of their choice sets. These correspond to a consumer's compensating variation for a change in product attributes (Small and

Rosen, 1981). The expected consumer surplus, CS_i, is defined as

$$CS_i = \frac{1}{|\alpha|} ln \sum_j e^{X_j \beta_{ij} - \alpha price_j}, \tag{19}$$

where α denotes the marginal utility of price. We estimate the consumer surplus for the choices as they are and the consumer surplus for the best alternative when no scores are available. We obtain changes in consumer surplus for each consumer, and then estimate the average change in consumer surplus as well as how changes in consumer surplus are related to consumer demographics by estimating the following equation:

$$\Delta(CS)_i = \delta_0 + \delta_1 D_i + \epsilon_i \tag{20}$$

where $\Delta(CS)_i$ is a vector of all the consumers' individually estimated changes in CS for the policy simulation of no score labels, D_i are the demographic characteristics (including the environmental score) of consumer i, and δ_0, δ_1 are parameters to be estimated.

5.3 Micro Level Results from Policy Simulation of Removing Scores

First, we estimate the predicted average probabilities of the choices and report the aggregate probabilities for each of the subgroups of products consumers can choose from, given the estimated parameters of column (2) in Table 6. There are four possible subgroups of products: purchasing one of the labeled products (type g=1), purchasing one of the unlabeled products that has no score (type g=2), purchasing one of the unlabeled products that has a score (g=3), and none of those options (g=4). These are depicted in the top left panel of

Figure 3 with the confidence intervals for inside options (g=1,2,3) and the outside option alternative (g=4).

At baseline, with scores displayed, the most chosen subgroup is the unlabeled wine, (g=2), then the labeled wines (g=1), followed by the outside option, (g=4), and finally, the unlabeled wines with scores, (g=3). When simulating the counterfactual choices of removing the score labels from the information set of the consumers, we will compare the panel on the right of Figure 3 with the left panel. For convenience, the change in shares is depicted in the bottom of Figure 3.

As shown in Figure 3, the average predicted probabilities change relative to the baseline. The most chosen option is not to select any of the inside options, and the lowest drop in purchases is for the labeled subgroup, while the unlabeled subgroup (g=2) share increases by less than the decrease in (g=1). Given that the outside option increases substantially, and its utility is normalized to zero, it is expected that the consumers that switch options have a lower utility than previously. We investigate formally the changes in consumers' surplus by comparing the baseline and the counterfactual scenarios compensated variation for all consumers.

We investigate whether there are significant heterogeneous changes in consumer surplus, by estimating equation (19). The estimates are reported in Table 9. On average, consumers lose from this policy experiment. Most choice changes are being driven by previous consumers of wine, who value scores, now choosing the outside option and purchasing no wine. The findings show that higher income is correlated with smaller consumer surplus loss given the positive and significant point estimate for income, and home ownership is uncorrelated with changes in CS, after we control for income. Consumers' age is uncorrelated with consumer surplus losses, given the insignificant coefficient associated with the increasing age of consumers. Finally, women lose less than men due to this policy experiment.

6 Conclusion

This paper estimates flexible discrete choice demand models to infer whether consumers place a significant value on the reduction of asymmetric information about wine quality, in the form of expert opinion scores. We use a macro and micro level dataset, consisting of retail scanner data at the time of a field experiment that reveals the expert opinion scores to retail consumers at the point of purchase.

We find robust results in terms of consumer valuations for expert scores, using both datasets. In particular, we obtain an implied average willingness to pay between 2 (using the macro level dataset for treated and control stores) and 3.2 dollars (using the micro level dataset for the treated store) for an average expert wine score of 83. We find that there is heterogeneity in WTP for wines originating from different regions and for different varietals. While using microdata, we overestimate the effect of expert scores as we cannot control for preferences in control stores like we do in the macro level data. However, we benefit from obtaining micro level data and the accompanying consumer characteristics to estimate observed heterogeneity in the effect of expert opinions. Indeed, we find significant heterogeneity, in that scores matter less the more likely a consumer is to own a home, or the higher his income. Other consumer attributes do not explain heterogeneity in preferences towards higher quality in the form of a score. Since homeowner status is correlated with income, our findings suggest there to be market potential in nudging lower income consumers who want to purchase better quality wines with expert scores.

Using counterfactual simulations, we estimate the changes in consumer surplus due to available quality information from expert opinion scores. Basing our welfare simulations on the micro level dataset, we find that removing scores leads to significant welfare losses for lower income consumers and for men. Furthermore, by using aggregate choices for treated and control stores in the macro dataset, we estimate there to be a significant loss in consumer surplus of 178 dollars, which represents 1.1 percent of the revenues in this market. This suggests that disclosing expert opinions results in small but significant positive welfare effects. Extrapolating to the national market, given total US wine retail revenues for 2013 were $36.3 billion dollars,[4] our findings would imply that consumers would be willing to pay up to 363 million dollars for expert opinions about the quality of wines.

7 References

Ali, H.H., S. Lecocq, and M. Visser, 2008. "The Impact of Gurus: Parker Grades and En Primeur Wine Prices." The Economic Journal, 118 (June): 158173.

Ackerberg, D., 2003. "Advertising, Learning, and Consumer Choice in Experience Good Markets: A Structural Empirical Examination." International Economic Review.

Baltas, G., 2001. "Expert Opinion Labeling: Issues and policies." European Journal of Marketing. 35(5/6):708-721.

Bazoche, P., C. Deola, and L. Soler, 2008. "An experimental study of wine consumers willingness to pay for environmental characteristics." 12th Congress of the European Associa-

[4] http://www.wineinstitute.org/resources/pressroom/04242014.

tion of Agriculture Economists. http://www.legrenelle-environment.fr/grenelleenvironment.

Berning, J., J.J. McCluskey, H.H. Chouinard, K. Manning, D. Sprott, and S.B. Villas-Boas, 2007. "Grocery Store Shelf Labels, Expert Opinion and Pricing Information, and Consumer Choice." mimeo.

Berry, S., J. Levinsohn, and A. Pakes. "Automobile Prices in Market Equilibrium." Econometrica 63(1995):841-890.

Bonnet, C., J. Hilger, and S. Villas Boas, 2016. "Reduced Form Evidence on Belief Updating under Asymmetric Information - The Case of Wine Expert Opinions." Working paper.

Bresnahan, T.F. "The Apple-Cinnamon Cheerios War: Valuing New Goods, Identifying Market Power, and Economic Measurement." The Economics of New Goods, (1997), edited by Bresnahan, T. F and R. J. Gordon, The University of Chicago Press, Chicago/London.

Burton, S. and J.C. Andrews, 1996. "Age, product expert opinion, and label format effects on consumer perceptions and product evaluations." The Journal of Consumer Affairs. 30(1):68-89.

Carson, T.S., N.E. Flores, K.M. Martin, and J.L. Wright, 1996. "Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods." Land Economics, 72(1), 80-99.

Caswell, J.A. and D.I. Padberg. "Toward a More Comprehensive Theory of Food Labels." American Journal of Agricultural Economics, 74(1992):460-68.

Dranove, D., D. Kessler, M. McClellan, and M. Satterthwaite, 2003. "Is More Information Better? The Effects of 'Report Cards' on Health Care Providers," Journal of Political Economy.

Drichoutis, A.C., P. Lazaridis, R.M. Nayga Jr., 2005. "Expert opinion knowledge and consumer use of expert opinion food labels." European Review of Agricultural Economics. 32(1):93-118.

Drichoutis, A.C., P. Lazaridis, R.M. Nayga Jr., 2006. "Consumers' use of expert opinional labels: a review of research studies and issues." Academy of Marketing Science Review. 9:1-22.

Frewer, L.J., C. Howard, D. Hedderley, and R. Shepherd. "Reactions to Information about Genetic Engineering: Impact of Source Characteristics, Perceived Personal Relevance and Persuasiveness." Public Understanding of Science 8(1999):35–50.

Golan, E., F. Kuchler, and L. Mitchell. "Economics of Food Labeling." Washington, DC: U.S. Department of Agriculture, Agricultural Economic Report 793, December 2000.

Hausman, J. A. and G.K. Leonard. "The Competitive Effects of a New Product Introduction: A Case Study." The Journal of Industrial Economics 3(2002): 237-263.

Hausman, J. A., 1997. "Valuation of New Goods under Perfect and Imperfect Competition." The Economics of New Goods. edited by Bresnahan, T. F and R. J. Gordon, The University of Chicago Press, Chicago/London.

Hensher, D., J. Louviere, and J. Swait, 1998. "Combining sources of preference data." Journal of Econometrics, 89 (12): 197221

Hilger, J., G. Rafert, and S. Villas Boas, 2011. "Expert Opinion and the Demand for Experience Goods: An Experimental Approach." Review of Economics and Statistics, 93 (4): 1289-1296.

Ippolito, P.A., and A.D. Mathios, 1993. "Information, Advertising and Health Choices: A Study of the Cereal Market." Rand Journal of Economics, 21(3):459-80.

Kiesel, K. D. Buschena, and V. Smith. “Do voluntary Biotechnology Labels Matter to the Consumer? Evidence from the Fluid Milk Market.” American Journal of Agricultural Economics 87(2005) : 378-392.

Kiesel, K., and S.B. Villas-Boas, 2007. “Got Organic Milk? Consumer Valuations of Milk Labels after the Implementation of the USDA Organic Seal.” Journal of Agricultural & Food Industrial Organization: Vol 5: No. 1, Art 4.

Kim, D. “Estimation of the effects of New Brands on Incumbent’s Profits and Consumer Welfare: The U.S. Processed Cheese Market Case.” Review of Industrial Organization, 25(2004):275-293.

Levy, A.S., S.B. Fein, and R.E. Schucker, 1991. “Expert Opinion Labeling Formats: Performance and Preference.” Food Technology, 45 (July), 116-121.

Loureiro, M., 2003. ”Rethinking new wines: implications of local and environmentally friendly labels.” Food Policy, (28:5-6), pp. 547-560.

Mathios, A.D., 2000. “The Impact of Mandatory Disclosure Laws on Product Choices: An Analysis of the Salad Dressing Market.” Journal of Law and Economics XLII (2000): 651-676.

McCluskey, J.J. and M.L. Loureiro, 2003. “Consumer Preferences and Willingness to Pay for Food Labeling: a Discussion of Empirical Studies.” Journal of Food Distribution Research, 34(3): 95-102.

McFadden, D. “The Measurement of Urban Travel Demand.’ Journal of Public Economics, 3(1974):303-28.

McFadden, D., 1981. “Structural Discrete Probability Models Derived from Theories of Choice. Structural Analysis of Discrete Data and Econometric Applications. C. Manski and

D. McFadden (eds), MIT Press: Cambridge, 198-272.

McFadden, D., 1999. "Computing Willingness-to-Pay in Random Utility Models." Trade, Theory, and Econometrics: Essays in Honour of John S. Chipman. J. Moore, R. Riezman, and J. Melvin (eds.), Routledge: London.

McFadden, D. and Train, K.E. "Mixed MNL Models for Discrete Response." Journal of Applied Econometrics 15(2000):447-470.

Nevo, A. "A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand." Journal of Economics & Management Strategy, 9(2000):513-548.

Nevo, A. "New Products, Quality Changes and Welfare Measures Computed from Estimated Demand Systems." Review of Economics and Statistics 85(2003):266-275.

Rao, A.R., 2005. " The Quality of Price as a Quality Cue." Journal of Marketing Research, XLII : 401-405

Reynaert, M. and F. Verboven, 2014. "Improving the performance of random coefficients demand models: the role of optimal instruments," Journal of Econometrics, 179: 83-98.

Roe, B. and M.F. Teisl., 2007. "Genetically Modified Food Labeling: The Impacts of Message and Messenger on Consumer Perceptions of Labels and Products.' Food Policy, 32(2007):49-66.

Sellers-Rubio, R. and J.L. Nicolau-Gonzalbez, 2016. "Estimating the willingness to pay for a sustainable wine using a Heckit model." Wine Economics and Policy, 5, 96-104.

Shiv, B., Z. Carmon, and D. Ariely, 2005. "Placebo Effects of Marketing Actions: Consumers May Get What They Pay For." Journal of Marketing Research, XLII : 383-393.

Small, K.A. and Rosen, H.S., 1981. "Applied Welfare Economics with Discrete Choice Models." Econometrica 49(1981):105-130.

Song, H., Z. Gao, and B. Gartner, 2015. "Willingness to pay for wines: A study of Hong Kong wine consumers." The 9th Annual Conference of the American Association of Wine Economists (AAWE), 26-30 May 2015. Mendoza, Argentina.

Train K.E., 2002. "Discrete Choice Methods with Simulation." Cambridge: University Press.

Whitehead, J.C., S.K. Pattanayak, G.L. Van Houtven, and B.R. Gelso, 2008. "Combining revealed and stated preference data to estimate the nonmarket value of ecological services: an assessment of the state of the science." Journal of Economic Surveys, 22: 872908.

Wine Institute, 2014. "California Wine Sales Grow 3% by Volume and 5% by Value in the U.S. in 2013." Apr 24, 2014, http://www.wineinstitute.org/resources/pressroom/04242014.

Figure 1: Kernel density of Scores of Treated Wines in Treated and Control Stores

Note: This figure displays jointly the kernel density estimates of the score distribution for the set of treated products in the treated store and the kernel density estimates of the score distribution in control stores for the same group of wine products treated in the treated store, given that we can see the same products in the control stores. The Kolmogorov-Smirnov (KS) test cannot reject the equality of treated wines scores' distributions in the treated and in the control stores, given that the KS test statistic is 0.0232 (p value 1.000).

Table 1: Summary Statistics of Wines for Treated and Control Stores

	(1) Treated Store Treated Wines	(2) Control Stores Treated Wines
Quantity (March)	16.99	11.24
	(26.01)	(26.02)
Quantity (April)	14.88	10.97
	(22.23)	(22.53)
Price (March)	10.98	11.24
	(5.00)	(5.03)
Price (April)	10.96	10.97
	(5.15)	(4.83)
% discounted (March)	0.91	0.88
% discounted (April)	0.88	0.89
Score	83.21	83.12
	(3.28)	(3.37)
% red	0.58	0.59
% white	0.35	0.34
Number of Wines	101	101
Number of Observations	2562	11055

Standard Deviations in parentheses. First column for Treated Store, next for Control stores.
Source: Scanner data set.

Table 2: Characteristics of Labeled and Unlabeled Products: Score, Varietal, Region, and Price

	Labeled	Unlabeled with Scores	Unlabeled Without Scores
Number of Observations	3004	42	18832
Average Score	83.69	86.69	N/A
Percentage Red	59%	14%	34%
Percentage White	37%	86%	62%
Percentage California	55%	62%	76%
Percentage Imported	45%	38%	24%
Average Price	9.56	24.47	9.21
Number Brands	17	5	62

This Table presents characteristics of wine products in the scanner data set of household purchases.

Figure 2: Trends of Market Shares of Treated Wines in Treated and Control Stores

Note: This figure displays jointly the evolution of the treated wine product market shares in the treated and control stores.

Table 3: Random Coefficient Logit Wine Demand Estimates for Treated Wines

	(1)	(2)
Score Level X Treated Store X Treated Period	0.005*	0.005*
	(0.003)	(0.003)
California	0.285***	0.192***
	(0.043)	(0.027)
Discount Dummy	0.921***	0.923***
	(0.029)	(0.025)
Treated Store X Treated Period	-0.334	-0.329*
	(0.206)	(0.197)
Treated Store	0.017	0.007
	(0.026)	(0.024)
Treated Period	-0.097**	-0.096**
	(0.042)	(0.040)
Price	-0.193	-0.199***
	(0.120)	(0.019)
Cabernet		0.357***
		(0.121)
Chardonnay		0.583***
		(0.114)
Merlot		0.134
		(0.115)
Pinot		0.224*
		(0.129)
Gris/ Gewurztminer		0.220*
		(0.118)
SD of Price	0.099*	0.100***
	(0.058)	(0.009)
Simulated GMM	3.146e-17	1.872e-18
Num of Obs.	13617	13617
Varietal FE		X

Standard errors in parentheses are clustered at the month level.

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table 4: WTP for Attributes of Treated Wines

	WTP
Score	0.025 *
	(0.015)
California	0.965 ***
	(0.136)
Cabernet	1.794 ***
	(0.608)
Chardonnay	2.930 ***
	(0.573)
Merlot	0.673
	(0.578)
Pinot	1.126 *
	(0.648)
Gris/ Gewurztminer	1.106 *
	(0.593)

Standard errors in parentheses. Estimates based on results from Table 3.
$*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table 5: Summary Statistics

PANEL A	California Population	Purchase Sample
Male	49.7	48.1
Female	50.3	51.9
17 or younger	24.4	0.0
18-59	59.3	30.2
60 or older	16.3	69.8
Less than some college	60.4	N/A
Associate degree, Bachelor degree	27.8	N/A
Graduate degree or more	11.8	N/A
$49,000 or less	41.5	4.8
$50,000-$99,999	28.9	14.8
$100,000 or more	29.4	80.4
White (Including Latino)	57.6	N/A
Black or African-American and minorities	42.4	N/A
Percentage likely to own a house	N/A	98.6
Number of Observations	38.8 million	3590

PANEL B	Sample With Demographics Information	Total Sample
Average Q Labeled Wines Pre Treatment (Pre)	1.330	1.329
Average Q Labeled Wines Post Treatment (Post)	1.502	1.446
Average Q Unlabeled Wines Pre	1.472	1.449
Average Q Unlabeled Wines Post	1.423	1.438
Share Labeled Pre	0.168	0.168
Share Labeled Post	0.175	0.166
Average Price chosen Pre	9.369	9.320
Average Price chosen Post	9.015	8.973
Percent Purchased on Discount Pre	0.942	0.930
Percent White Pre	0.592	0.587
Percent Red Pre	0.368	0.372
Percent California Pre	0.735	0.738
Number of households	3590	4754
Number of Observations	24610	31361

Source for California Population Data : 2014 CA Census Fact Finder Database.

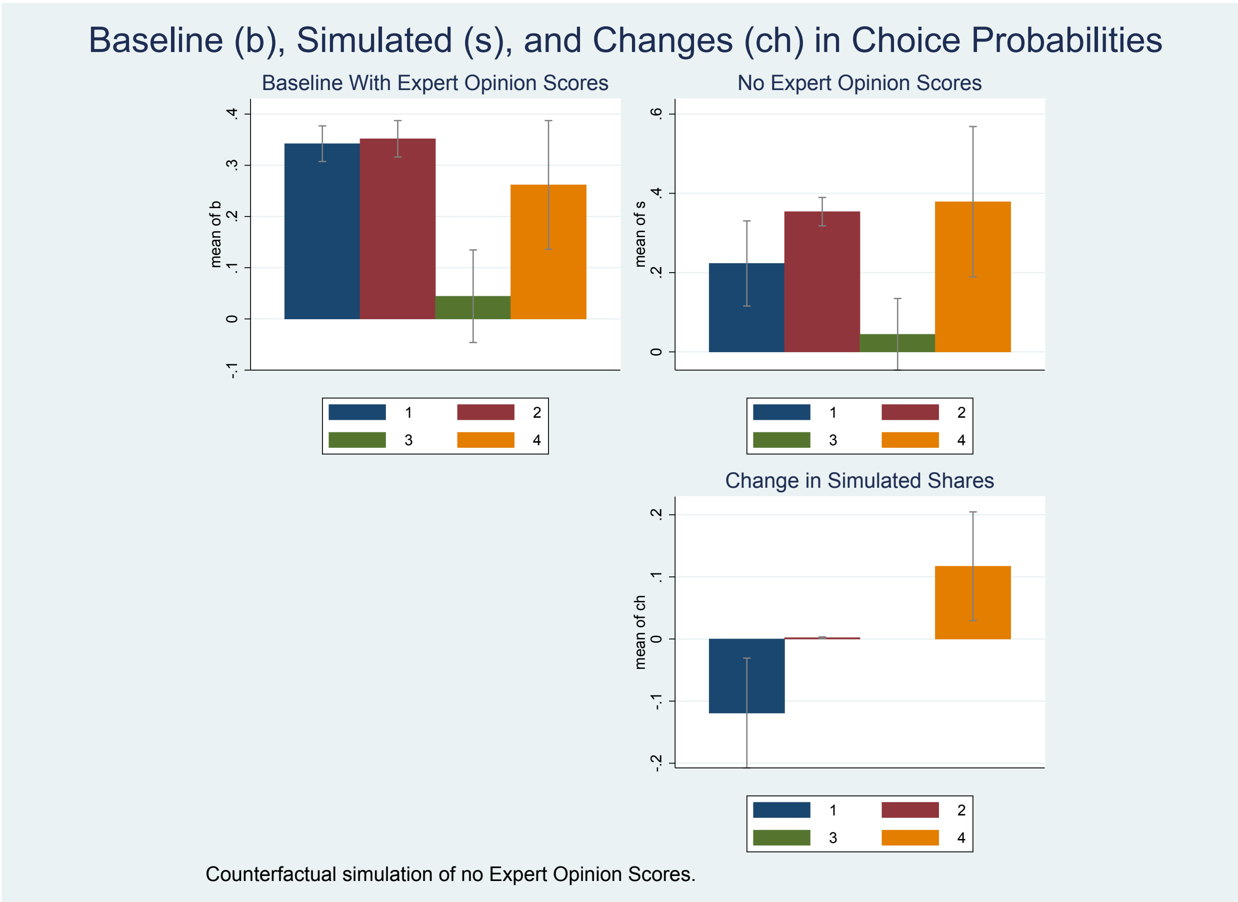
Table 6: Heterogeneous Logit Choice Estimates by Individuals Demographics

	(1)	(2)	(3)	(4)
price	-0.014***	-0.014***	-0.014***	-0.014***
	(0.003)	(0.004)	(0.004)	(0.004)
Treated Wine X Treatment Week = Treatment	-0.639	-1.552	-1.252	-1.942
	(1.386)	(1.432)	(1.406)	(1.449)
Treatment X Score	0.011	0.046**	0.040*	0.050**
	(0.018)	(0.022)	(0.021)	(0.023)
Treatment X Score X Income	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Treatment X Score X Age	-0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Treatment X Score X Female	0.000	-0.001	-0.002	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)
California	-0.162	-0.179	-0.180	-0.181
	(0.687)	(0.692)	(0.689)	(0.692)
red	0.016	0.008	0.020	0.008
	(0.067)	(0.069)	(0.067)	(0.069)
white	0.118*	0.117*	0.122*	0.117*
	(0.066)	(0.068)	(0.066)	(0.068)
Treatment X Score X High Likelihood Own House		-0.028***		-0.027**
		(0.010)		(0.010)
Treatment X Score X Likelihood of Owning House			-0.003***	
			(0.001)	
Treatment X Score X California				0.003*
				(0.001)
Constant	-1.639**	-1.639**	-1.650**	-1.634**
	(0.684)	(0.687)	(0.684)	(0.687)
Num of Obs.	10312	10052	10312	10052
R squared	0.220	0.221	0.220	0.222
Week FE	Yes	Yes	Yes	Yes
Brand FE	Yes	Yes	Yes	Yes

Robust standard errors in parentheses. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Note: The table displays the estimates of Logit regressions where the dependent variable is equal to log of share of an alternative chosen minus the log of the share of the outside option.

Figure 3: Estimated Households' Choice Probabilities at Baseline and After Simulation of no Scores



Note: The figure displays the estimated baseline probabilities of choosing a Labeled product (g=1), an unlabeled and unscored product (g=2), an unlabeled but scored product (g=3), and the outside option (g=4).

Table 7: Random Coefficient Logit Estimates of Wine Demand

	(1)	(2)
Treated Wine X Treatment Week (= Treatment)	0.556	-1.727
	(1.672)	(1.317)
Treatment X Score	-0.006	0.054**
	(0.020)	(0.021)
Price	-0.014	-0.018***
	(0.043)	(0.001)
Treatment X Score X Income		-0.000
		(0.000)
Treatment X Score X Age		0.000
		(0.000)
Treatment X Score X Female		-0.003
		(0.002)
Treatment X Score X Likelihood Own House		-0.034***
		(0.010)
California		0.026
		(0.016)
Red		-0.003
		(0.035)
White	0.130***	0.114***
	(0.035)	(0.034)
SD of price	0.0259	0.0258
	(0.0659)	(0.0659)
Num of Obs.	21865	10176

Note: The table displays the estimates of Random Coefficient Mixed Logit regressions where the dependent variable is equal to log of share of an alternative chosen minus the log of the share of the outside option. Standard errors in parentheses. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table 8: Regression of Respondents' Mixed Logit WTP Estimates on Demographics

	(1) Heterogeneous Conditional Logit WTP for Scores
Household Income	-0.012*
	(0.006)
Age of he head of household	-0.071
	(0.055)
Female	-1.715
	(1.507)
Likely to Own a Home=1, Unlikely=0	-1.164
	(4.717)
Constant	13.167**
	(5.516)
Num of Obs.	1998
R squared	0.002

Robust standard errors in parentheses. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$

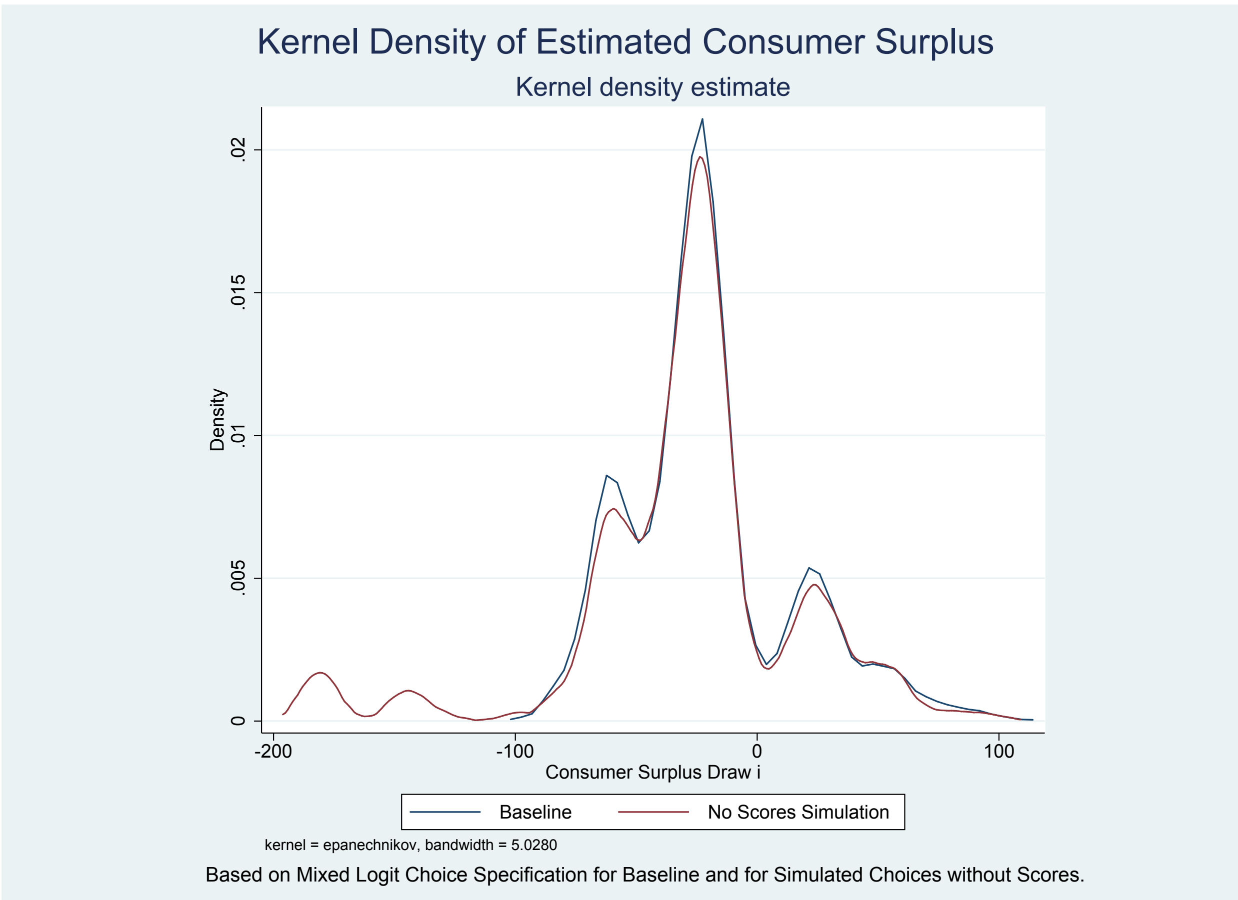
Table 9: Regression of Change in Consumer Surplus Estimates on Demographics

	(1) Dependent Variable Change in CS
Age of the head of household	0.078
	(0.087)
Female	4.934**
	(2.206)
Likelihood of household Cardoholder Owning a Home	6.085
	(7.923)
Household Income	0.027***
	(0.010)
Constant	-25.012***
	(9.154)
Num of Obs.	1215
R squared	0.009

Change in consumer Estimates from Simulation of Removing Score Labels.

Standard errors in parentheses. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Figure 4: Distribution of Estimated Consumer Surplus With and Without Expert Opinion Scores- Household Level Panel Data Analysis



Note:This Figure displays jointly the kernel density estimates of consumer surplus in the baseline with scores and in the counterfactual scenario without scores. The Kolmogorov Smirnov test for equality of both distributions is rejected. All estimates are based on the demand estimates in Table 6.