Reduced Form Evidence on Belief Updating under Asymmetric Information - Consumers' Response to Wine Expert Opinions

Céline Bonnet, James Hilger, and Sofia B. Villas-Boas*

January 17, 2020

Abstract

We estimate the effect of quality labels on purchases through a retail field experiment. Utilizing product-level panel scanner and product characteristic data for both labeled and unlabeled wines we estimate the average and heterogeneous effect on purchases. Consistent with earlier work, we find an average effect that is positively correlated with scores. We advance the consumer belief and product information literature on two fronts. One, higher scores matter more for lower priced products. Two, spillover effects impact sales of untreated wines; these effects can be positive or negative and are impacted by the average score and label converge within brand.

Keywords: Field experiment, labels, information, expert opinion, wine, product attributes.

JEL Classification: C23, C25, D12, H20.

1 Introduction

There is a relatively large theoretical literature that discusses the ways in which informed consumers and expert opinion may solve adverse selection problems (Akerlof, 1970, Bagwell and Riordan 1991). Despite the growing body of empirical work that has applied this theory

*We thank the editor, Carl Johan Lagerkvist, and three anonymous referees for their comments. Bonnet: Toulouse School of Economics, INRA, University of Toulouse I Capitole, Toulouse, France; Hilger: U.S. National Marine Fisheries Service; Villas-Boas: Department of Agricultural and Resource Economics, University of California, Berkeley, CA, 94720. We thank 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 and thank the United States Department of Agriculture, specifically the National Institute of Food and Agriculture for support. We thank Peter Berck for comments. Corresponding author: Céline Bonnet, celine.bonnet@tse-fr.eu.

to data (Hilger et al, 2011, Sorensen and Rasmussen, 2004, Reinstein and Snyder, 2005), there remains an empirical gap. There is little research in which consumers' actual choices are used to test the theoretical mechanisms that underlie the ways in which consumers make inferences about the quality of products with and without expert opinion information. This is particularly true for products where producers use observable attributes such as price and brand advertising to signal quality. The contribution of this paper is to test for some economic mechanisms that explain consumers' beliefs and to find evidence on how consumers use information about observable characteristics to infer the quality of products.

To analyze this question, we use purchase decisions from a retail field experiment in which we label a random subset of scored wine products in one treated store and do not label those same scored products in comparable control stores. Previous studies suffer from the fact that experts tend to give high rankings to high-quality products, so it is difficult to examine the extent to which demand is affected by the quality itself versus the opinion. As in Reinstein and Snyder (2005) or Hilger et al. (2011), we circumvent this obstacle using an experiment where a random subset of wines with scores receive an expert opinion label on the shelf, a subset of wines with scores do not, and consumer also see wines without scores on the shelves in their choice set. The scores are given by a proprietary source and the scoring system is based on a variety of expert wine tastings. We combine a product weekly scanner data set with detailed product-specific attribute information readily available to all consumers, such as brand, variety, region of production, and the price. Using revealed consumer preferences, the objective of this paper is first to estimate the average and heterogeneous effects of the field experiment on consumption of three categories of wine. "Treated" wine has received a quality score from wine experts; the treatment consists of putting a label on the bottle that shows the score, in addition to the usual wine bottle label and the price sticker. "Unlabeled and scored" products have received an expert score, but the score is not shown on the bottle. "Unscored" products have not been scored, so they do not have the treatment label. Moreover, we investigate heterogeneous treatment effects according to product characteristics such as price range or regions of production. This analysis also tests for evidence of spillover effects among products of the same brands or same varieties, to reveal possible mechanisms behind how consumers make inferences about the quality of products they buy, given observable product attributes in the choice set.

This paper uses the same experimental design as Hilger et al. (2011), but extends the analysis to purchase data of all three categories: treated (scored and labeled) wines, unlabeled wines with scores, and unlabeled wines without scores.[1] We match the treatment store to control stores, making sure all three subgroups of wines (treated, unlabeled with scores, and unlabeled wines without scores) share common pre-period trends in the treatment and control stores. Our contribution is to extend Hilger et al. (2011) to an investigation of incomplete information through testing the possible mechanisms behind how consumers form beliefs about product quality when making their revealed choices given available information. While Hilger et al. (2011) documented a carefully identified causal positive average effect of expert opinions on wine demand, especially for higher scores, and no average effect on the unlabeled wines, we dig deeper in this paper understand why this is so, and how consumers incorporate the new information in deciding what future choices to make. For instance, we test whether there are positive or negative spillover effects on sales of unlabeled products

[1] In Hilger at al. (2011), the analysis mainly considered treated wines and only a robustness check was made on untreated wines.

that belong to the same brand or variety as labeled products. We also investigate whether such spillover effects vary with price.

The remainder of the paper is structured as follows. Section 2 discusses the literature of consumer belief. Section 3 presents the experimental setting and the data. Section 4 describes the empirical strategy to estimate the effects of displaying labels on average and by product attributes to test theories of belief formation in the absence of quality information. Section 5 presents and discusses the findings and section 6 concludes.

2 Background literature on consumers' beliefs

The quality of food products can be difficult to observe when consumers make their choices. Their choices are then based on quality expectations. There is a large empirical literature on understanding consumer beliefs about product quality. In the drink and food markets, the perception of quality can be measured using two different scales: one for intrinsic attributes and one for extrinsic attributes (Jover et al., 2004). However, in real choice situations, only the extrinsic characteristics of the food products can be used to infer the quality of products. Among extrinsic characteristics, consumers may consider the price, brand name, and country of origin, as well as promotional activities associated with the products in their choice set (Ozretic-Dosen et al., 2007; Schnettler et al., 2009 for beef; Grankvist and Bien, 2007 for organic food products).

In the wine market, the experimental studies show that price, country of origin and expert scores affect quality expectations (Mastrobuoni, et al, 2014; Gneezy at al., 2014; Veale and Quester, 2009; Palma et al, 2013; Chocarro and Cortiñas, 2013). Some theoretical work has investigated the mechanisms of the quality expectations / product characteristics relationship. Indeed, price and brand advertising can be used by firms to signal quality in a context of asymmetric information and repeated interactions. Grossman (1981) show that firms have an incentive to communicate their quality to reduce the assymetry of information. Shapiro (1983) and Bagwell and Riordan (1991) find that high prices signal quality if consumers attribute high quality to seeing a high price and that there is a proportion of informed consumers who buy if they believe that quality is good.[2] If that is so, revealing high scores for high-priced products will have less effect than revealing high scores for lower priced similar products. However, if a low price signals high quality through promotions to induce trial and repeated purchases once quality is assessed, as in Mahenc (2004), Spence (1976) and Tellis and Wernerfelt (1987), then high scores for high-priced products may have a larger effect on demand than high scores for low-priced products, which consumers already inferred were good quality in prior to the additional label information).

The objective of this paper is to empirically test theories of equilibrium price and quantity to learn about quality expectations of consumers. In the context of asymmetric information and a choice set that has wine products with different combinations of product attributes, consumers infer product quality based on the combination of those attributes. Wine is an experience good, in that consumers only realize its quality after consumption, and sometimes

[2]Shapiro (1983) found that adverse selection is prevented through an incentive-compatible equilibrium price-quality schedule, which prevents the sale of low-quality goods at a high price (which would imitate good quality) and sells high-quality products at a premium above their cost, so that this premium compensates the cost of providing the high quality product to all informed consumers and to a proportion of randomly purchasing uninformed consumers. Bagwell and Riordan (1991) show that the high quality producers' price decreases as more consumers are informed.

even after consumption, may not fully appreciate its quality (Ali and Nauges, 2007; Veale, 2008; Scozzafava et al., 2016). There may be a role for experts to reveal their opinion on wine product quality in order to help consumers make choices given the choice set of wine products with different prices, attributes, and the additional expert scores. In this context of revealed quality information through the expert score, we investigate whether the change in quantity purchased occurs among low or high priced wines, and test if quality is initially signaled through low or high prices. Moreover, revealing information on a product could also lead to spillover effects on other products. Indeed, product brand information could improve the reputation of other products within the same brand and such spillover effects could vary with prices. Costanigro et al. (2010) show that there exists a relationship between price and the structure of the name (collective and specific) and reputation in the wine market.

Contrary to the large empirical literature on understanding consumer beliefs about product quality, which is based on experimental analyses, we will use consumers' purchases in stores that reveal the true preferences in a real context of purchase. Indeed, 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).[3] We follow the empirical literature that uses a difference and difference approach to infer the causal effect of expert opinions on demand using real

[3]Consumers consider the internal costs and benefits of their potential choices and experience the consequences of their actions. Carson et al. (1996) shows through meta-analysis that estimates from stated and revealed preferences differ. Empirical literature using revealed choices 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, 2007), and advertising (Ackerberg, 2003).

consumer purchases. Hilger et al. (2011) on the wine market, Reinstein and Snyder (2005) for the movie industry, and Friberg and Gronqvist (2012) for the wine market identify the effect of expert opinion on demand but without explaining why. Our contribution is to use this empirical method to identify the previously described mechanisms of consumer belief about wine quality.

3 The experimental setting and the data

To examine consumer demand responses from randomly disclosing expert opinion only on a subset of high quality wines as well of spillover effects into lower quality (unscored) wines, we use an experimental design. Wine scores from a proprietary wine scoring system were displayed in the treatment store for four weeks during the month of April 2006.[4]. We use a proprietary score wine data set to randomly label 150 wine products with score labels that were attached below the price tag every week by the research team. The scoring system in place utilized a 100 point scale that is commonly utilized in other rating systems.

Among the wines with scores, the division of wines between those labeled and those not labeled is based on a random assignment decided by the authors. As a consequence, not all wines with scores received a label. More specifically, 325 scored wines remained unlabeled. In addition, there remained nearly 613 wines in the consumers' choice set that were not scored by the experts. Thus the consumers were presented with three subgroups of wines:

[4] Although the data are not recent, they are unique in that they pertain to a unique and first time implemented field experiment in the retail setting. The insights are from consumer actual choices rather than survey evidence or lab experiments, and we got access to these data because the retailer did not deem them no longer an issue to be analyzed

1) labeled wines with scores (14% of the choice set), 2) unlabeled wines with scores (28% of the choice set), and 3) unlabeled wines without scores (58% of the choice set).

Each label features the name of the proprietary scoring system and the wine's score that. In theory, scores range from the lowest (50) to the highest (100), but scores below 70 are not released by the rating agency. All wine bottles remain on the shelves in their usual shelf display location, so consumers are presented with the same assortment of wines on the shelves before and after the labeling experiment. Because shelf display and product assortment remains unchanged during the labeling experiment, consumers see two sets of wine products in the treated store: labeled and unlabeled. Given that unscored wine products may be of lower quality, displaying only a random subset of scored wine products in the labeling experiment allows us to estimate spillover effects within product characteristics (such as brand or variety) into unlabeled wines with scores, as well as into unscored wine products. In so doing, we infer whether consumers update their beliefs about product quality of unlabeled wines. Figure 1 displays the kernel density of the score histogram for labeled and unlabeled wines, with scores in the treated store on the left panel and for the control stores on the right panel. Using a Kolmogorov-Smirnov (KS) test for equality of the distributions for both the labeled and unlabeled panels, we find a KS estimate of 0.0987 for the Treated Store with a p-value of 0.481 and a KS estimate of 0.0935 with a p-value of 0.518 for the Control Stores; therefore, we cannot reject the equality of the score distribution between the labeled and unlabeled status of wines. There is therefore a nice match in the whole distribution of labeled and unlabeled wines in the treatment store and in the control stores, which supports the difference in differences estimation strategy for the identification of the effect of score label.

The treatment store is located in a wealthy suburban neighborhood and is in the same marketing division as a set of 38 potential control stores in Northern California.[5] The shared marketing division means that the pricing, promotion, and display layout is common among the treatment and control stores. This contributes to a good balance of observable determinants of quantities of wine sold that originate from the retail marketing strategy.

3.1 Wine purchase data

The data set consists of records for all wine products purchased in a certain week by store, from March 2003 until May 2006, which we will aggregate to the product by month by store level for our analysis. In particular, we observe purchase data by store and by week for 102 products among the 150 wine products for which we attached labels to the price tag every week. For the sub-sample of 325 scored wines that remained unlabeled, we observe purchase records for 230 products. In addition, we have purchase records for 610 of the 613 wines in the choice set of consumers that were not scored by the experts. Overall, we have purchase data for all wine products in the treatment store and a set of potential control stores, among which we will use only the four that best match the treatment store's pre-period trends in labeled, unlabeled, and unscored wines, as we will show next.[6] The scanner data provide a unique wine product code identifier (UPC), the brand name and variety of the wine, the

[5] As in Hilger et al, (2011) on the one hand, if consumers in wealthy areas are likely to be more fully informed regarding wine quality than consumers in other areas, one could argue that we have a low likelihood of finding a significant treatment effect. On the other hand, if consumers in wealthier areas care more about quality and expert reviews, we would have a higher likelihood of finding an effect.

[6] We use past data to match the treated store and possible controls based on similar pre trends not only for the labeled products but also for unlabeled scored products and unlabeled un-scored products unlike in Hilger at al (2011) where more control stores are used.

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 unique product code and store to generate the number of bottles, average shelf price, average price paid (the shelf price net of discounts), and whether a bottle of wine was discounted. Pricing and discounting for each product are common for all the stores in the data and, moreover, wine pricing was not updated due to the selection of products into labeled and unlabeled status. Lastly, prices were not differentially updated in the treated store due to our experiment.[7]

For those wines for which proprietary wine score data exist, we merge the wine score data with the scanner data. Detailed product attribute data includes the brand of the wine product, variety (for example Cabernet or Merlot) and aggregate wine type (red, white, or other). The origin of production and imported status are merged with the scanner data set.

We use data for multiple months before the treatment window to investigate pre-period trends. As reported in Table 1, in the treatment store, we have data on 2,562 product month observations for treated wine products, 5,721 observations for untreated wine products with scores, and 16,194 observations for unscored wine products. In the four control stores, the total number of observations for the three treatment status subgroups of wine products are 11,058, 24,578, and 72,871 respectively. The three product subgroups have the same proportion of observations across the treatment and control stores. The three sets of products are in the consumer choice set.

Sample descriptive statistics are reported in Table 1 and organized by store and wine treat-

[7] While in our empirical setting the price does not change due to experimental labels, it is worth noting that an extensive stream of the literature is related to the effect of experts opinion on wine prices (Dubois and Nuages, 2010; Hadj Ali et al., 2010)

ment group as follows. In the first three columns, we report descriptive statistics in the treatment store for treated/labeled wines in column 1, unlabeled wines with scores in column 2, and untreated wines without scores in column 3. Columns 4, 5, and 6 report descriptive statistics for the control stores for the respective three wine and score groups.

In the first row, we report average quantity sold during the pre-treatment month (March); this is followed by its standard deviation in the second row. Below that is the average quantity sold during the treatment month (April), followed by its standard deviation. Average quantity sold by product in the pre-treatment month is 17 bottles for labeled wines in the treatment store and 19.2 bottles in the control stores. Unlabeled wines with scores sell 25.6 bottles on average in the pre-treatment month in the treated store and 29.1 bottles in the control stores on average. Finally, unscored wines sell about 12 bottles by product in the treated store and 13.4 bottles by product in the control stores during the pre-treatment month. While the unscored wines sell on average fewer bottles per month than the wines with scores (both labeled and unlabeled), this pattern is consistent in the treated and control stores. Given our panel data, we can control for differences in average bottles sold, by a fixed effect in our specification, and obtain our estimates of interest on the residual variation in quantities sold after controlling for constant product-specific bottles of wine sold.

Average prices are not statistically different between the pre-treatment and treatment month; this is the case regardless of treatment status. As can be seen in Table 1, the prices of scored (both labeled and unlabeled) and unscored wines are quite similar to each other, the average prices for all the three sets of products in consumer choice sets is 10.98 dollars for scored and labeled, 10.15 dollars for unlabeled wines with scores and 10.98 for wines without any scores. Most of the wine consumers purchase is discounted in the both the pre-treatment and treatment months: approximately 90 percent for treated wines (in both treatment and control stores), almost one hundred percent for untreated wines with scores, and approximately 75 percent for untreated wines without scores. Given that prices are on average similar across the three groups of products, we can assume that the relationship between price and quality expectations of consumers does not differ on average across the three groups.

The next two rows of Table 1 reports the average and standard deviation of scores by treatment status for treated and control stores. Here we see that average scores at the treatment store are not statistically different between the treated wines and untreated wines, with scores of approximately 83 and standard deviations of roughly 3.2. This is corroborated by Figure 1's reporting of similar score distributions among treatment status for the treatment and the control stores.

The next two rows report the composition of red and white wines for each store and wine treatment group pair. For wines within the treatment store, the treated group consists of 58 percent red wines, the untreated with score group consists of 45 percent, and the unscored group contains 50 percent red wine; a similar distribution characterizes the control stores. Similarly, white wine proportions by treatment status are similar between the treatment store and control stores.

There is comparable regional composition of the set of labeled wines (set 1), the wines with scores but no labels (set 2), and the set of unscored wines (set 3). For instance, with respect to the regional variation of the origins of the 150 treated wines, 85 percent are domestic (120 wines) and the remaining are imported: 8.5 percent are from Australia (12 wines), 2.9 percent from Italy (four wines), 2.1 percent from France (three wines), and the rest are two from New Zealand, one from Chile, one from South Africa. For the scored but not labeled set, 85 percent are domestic, and the remainder are imported: 8 percent from Australia, three percent from New Zealand, two percent from France, one percent from Italy and one wine from Spain. Finally, for set 3, the wines that receive no scores are 81 percent of domestic origin, and the remaining are imported: 6.5 percent from Italy, 5.5 percent from Australia, two percent from Spain, and less than one percent each from Germany, Argentina, Chile, Israel, South Africa, and New Zealand. The results are to be interpreted given this set of wines in terms of average effects and spillover effects being estimated among wines mostly originating from domestic producers, and the main imports being from Australia and to a smaller extent from Europe (Italy and France). Furthermore, among the treated domestic wines, there are a total of 20 regions of origin, where only 2.6 percent originate from Napa, and 2.6 percent originate from Sonoma.

3.2 Pre-period trends in treatment and control stores

To estimate the causal effect of revealing score information on a random subset of wines with scores on the three different subgroups of wine products as defined by treatment status, we need to verify that one crucial assumption is satisfied - that there are similar pre-period trends for the treatment and the control stores used in the analysis.

First we investigate whether the pre-period is balanced in terms of pre-existing trends in total quantity sold of wine products for the treated store and control stores. Figure 2 presents the total quantity sold by store per month for the year preceding the treatment, where the treatment is indicated by a vertical line. While the various stores differ in levels of quantity sold, their trends are quite similar. While the treatment store has a lower total quantity of wine sold than each control store, the trends for wine quantity purchases follow similar patterns in the treatment store and the four control stores.[8] Thus, while the treatment store has different quantities sold than the control stores, to the extent that these differences are constant over time, store fixed effects will control for all possible time invariant determinants of wine demand at the store level.

We repeat in Figure 2 the graphical analysis for the treated (labeled) products only. Once again, while there is a lower quantity of wine products sold in the treatment store than in each control store, the pre-treatment trends follow each other closely. This pattern is also evident for untreated wines with scores and for the un-scored wines, respectively. In sum, for all the subgroups, the trends are quite similar in the treated and control store, which will allow us to investigate the causal effects of the treatment on treated wines, untreated wines with scores, and untreated wines without scores.

4 Reduced form empirical specification

Using two data sets, a store level and a product characteristic data set, we estimate a reduced form revealed preference specification of consumer responses to the labeling treatment. In particular, we extend Hilger et al. (2011) by measuring heterogeneous changes in wine consumption depending on wine characteristics to test theoretical predictions of consumer inferences about product quality, by taking advantage of the exogenous changes in information

[8] As a more rigorous test of parallel trends, we regress quantity on a time trend for the treatment and control stores separately. We find that the point estimates of the trend coefficients in treatment and control stores are not statistically different from each other. Furthermore, the time series correlation of the sample averages of the treated and the control stores is high, suggesting that the treatment and control stores share broadly similar time varying patterns in the pre-treatment period.

about product attributes introduced through the field experiment.

4.1 Global identification strategy

We follow a difference-in differences-approach (DID) commonly used in the policy evaluation literature (see Meyer 1995; Bertrand, Duflo, and Mullainathan 2004). Our control structure is twofold: temporal, as we compare one particular product's purchases in weeks with and without labels; and cross-sectional, as we compare purchases of similar types of products between the treatment and the selected control stores. We look at wine products in the three different subgroups defined by treatment status. That is, we look at the labeled products with scores, the unlabeled products with scores, and finally the unlabeled products with no scores.

Let a product i be defined as a wine option in a certain store. The DID model specification is as follows :

$$\begin{aligned} Q_{igm} &= \alpha_i + \beta_1 \text{Price}_{igm} + \beta_2 \text{Treated Store}_{ig} + \beta_3 \text{Treatment Period}_m + \\ &+ \beta_4 (\text{Treated Store*Treated Period})_{igm} + \beta_5 \text{D}_{igm} + \epsilon_{igm} \end{aligned} \tag{1}$$

where Q_{igm} is the quantity sold of product i in product treatment status group g and month m, α_i is a product fixed effect, $Price_{igm}$ is the unit price of product i in product group g and month m, D_{igm} is a discount dummy if the product i in product group g and month m is on sale and Treated Store_{ig} is an indicator that product i in treatment status group g is in the treated store. A time indicator Treatment Period_m is equal to zero in the pre-treatment period and equal to one in the treatment period.

The coefficient of interest is the interaction of Treated $Store_{ig}$ and the Treatment $Period_m$.

When estimating equation (1) separately for g =(Treated Wines, Untreated Wines with Scores, Un-scored Wines), we obtain three coefficients of β_4, which are the average effect of the treatment on the wine demand of each of the subgroups of products defined by treatment status.

4.2 Heterogeneity and tests for belief updating

To investigate the heterogeneity of the effects according to observable product attributes, we split data into four subsets in the treated status: high score quartiles and low price quartile, high score quartiles and high price quartile, low score quartile and low price quartile, and finally low score quartile and high price quartile. Score range allows us to test whether revealing higher scores is associated with larger demand responses. We define a low score to be in the first quartile (25% percentile), which corresponds to a score lower than 81, and a high score to be higher than or equal to 81.[9] Price range allows testing for pre treatment consumer beliefs about wine quality. We then estimate the equation of the difference in difference specification for the four data subsets of wines separately.

If the estimated β_4 for the high score low price subset is positive and significant, then we reject the idea that, pre-treatment, consumers believed that a low price signaled high quality (as in Mahenc, 2004, Spence, 1976, and Tellis and Wernerfelt, 1987). However, if β_4 is not significantly different from zero for the high score, high price wines, this would be evidence consistent with Shapiro (1983) and Bagwell and Riordan (1991) that, similar to their pre-treatment beliefs, consumers believe that all else equal a high price means high

[9] A score equal to 81 was found to have a treatment effect that is equal to zero, based on the estimates presented in column (1) of Table 5 of Hilger et al. (2011) for labeled wines with scores only.

quality. This also would be true if β_4 is not significantly different from zero for low score - low price wines. We do not have enough observations in the high price - low score set to perform an empirical test for that group.

4.3 Test for spillover effects

When scores are revealed while holding constant other attributes, such as price (pre- and post treatment), brand, and variety, consumers may update the way they use price in conjunction with other attributes as a quality signal in the post-treatment period. We investigate whether there are significant spillover effects due to belief updating, by observing consumers' revealed preferences for unlabeled wines that share similar characteristics with labeled scored wines. In this last specification, we test for spillover effects by taking advantage of having collected variety and brand name attribute data for all wines. We also investigate whether such spillover effects vary with price.

In particular, for each wine, we compute variables that measure the proportion of wines that are treated by brand and variety, respectively. We also compute a variable defined as the average score of treated wines in the same brand of each wine, and a variable defined as the average score of treated wines in the same variety. To test for spillovers within brands, we test whether or not the degree of treatment, defined as the average score of treated wines within the brand or the number of treated wines within the brand, increases demand for wines in the same brand. If the coefficient is significantly different from zero, we reject the null of no spillovers. We also perform the same investigation to test for spillovers within varieties and compare the estimates of brand and variety spillovers. Lastly we investigate whether there are differential spillovers depending on wine prices as in Costanigro et al.

(2010).

5 Results

This section is organized as follows. First we present the average treatment effects on the treated wines. Then we turn to presenting heterogeneity estimates by score and by price range and testing belief updating about quality of wines. Finally, we look at spillovers among the treated wines and investigate whether there are spillover effects into unlabeled scored wines as well as un-scored wines of different price quartiles.

5.1 Average treatment effects on treated wines

We present the results from the reduced form specification of equation (1) in Table 2, where the dependent variable in the first two columns is the quantity of bottles sold of wine product i at store s and month-of-sample m and the ln of quantity in column (3). The first row contains the estimated treatment effects, which is the interaction of "Treated Store and Treated Period" dummy variables. We also include "Treatment Store" and "Treatment Period" dummy variables in the regression. In column (2), we report on a specification with the addition of a price variable and a dummy variable that indicates whether or not the product is discounted, while Column (3) repeats the specification in column (2) with quantities and prices in natural logarithms.

Treated products sell an average of 15 bottles a month. Between the pre-treatment period and the treatment period, average demand drops for both the treatment and the control stores, as shown by the negative albeit not significant coefficient of -3.277 in levels in column (2), which corresponds to a 19.3 percent decline given in (3). As already shown in Figure 2, average quantity is lower for the treated store than for the controls, which is reiterated by the negative and significant point estimate of "Treated Store", by about three bottles, or 18.6 percent. The price coefficient is statistically different from zero and negative in all specifications. Given the logarithm specification in column (3), we can interpret the coefficient on *Price* as the average elasticity of demand for a wine product keeping all other prices constant. This coefficient equals -2.671, which ultimately means that demand for wine is elastic. This estimate corresponds to the average price elasticity of demand for a wine product, keeping all other wine prices constant; as profit maximizing firms, wine producers should price on the elastic portion of the demand. Elasticity estimates are sensitive to the data used and also to the level of aggregation of products (see e.g., Rum et al. 2012, for beer elasticities in the US, ranging from -0.23 to -4.6). Our estimate is on the high side of the literature range (for instance in the UK estimates range from -0.14 for very aggregate demand models (Duffy, 2002) to -2.42 for more disaggregated ones (Crooks, 1989). However, considering the many possible substitutes in our choice set, and despite product differentiation via brands and varieties, the own-price elasticities should be quite large in absolute value. Our results are consistent with the analysis of Davis et al. (2008), who use a random utility approach to model the consumption behavior for differentiated brands within the wine market. Indeed, they find that own-price elasticities for demand of differentiated wine products range from -0.61 to -6.28, with an average around -2.

In all three columns of Table 2, the coefficients on the interaction terms "Treated Store X Treated Period" are positive. For the specification in levels, we see that revealing scores increased wine demand by 1.5 bottles, about five percent, and the increase is statistically different from zero, suggesting that wine product sales decreased less in the treated store in the treated period than in the control stores, leading to a positive average treatment effect. For the ln-ln specification, the coefficient is 0.052, but not statistically different from zero.[10] Next, we show heterogeneity by score levels and along other observable attributes for different subsets of wines according to treatment status and use a log specification for easier interpretation of the estimates as percent changes.

5.2 Treatment effects by score levels on treated wines

To further understand whether and how consumers change demand for different types of wine, we estimate equation (1) by interacting the treatment effect with the displayed level of score and present the estimates in column (1) of Table 3. In column (2) we estimate (1) for the subset of scores in the lowest score quartile, that is scores below 81, and in column (3) for wines with scores 81 and higher. Recall that we define a high score to be greater than or equal to 81 and a low score to be lower then 81, which is in the lowest score quartile. Price range and score range classifications allow testing for pre-treatment consumer beliefs about wine quality. We present here the estimates of equation (1) for the four data subsets, namely for the wines in the set of low price and low quality, the set of low price and high quality, the set of high price and low quality, and finally for the set of wines of high price and high quality. Breaking up the sample into the four sets is a flexible form that allows different (non-linear) effects to be estimated for price and ratings combinations without assuming a

[10] Our results are consistent with those of Hilger et al (2011) but the point estimates are not the same because we use a different set of control stores to estimate the average treatment effect here. In particular, we match the treatment store to the best controls in terms of pre-period trends for not only the treated wines but also for the untreated wines with and without scores, and use four control stores.

specific linear or quadratic functional form.[11]

We find that the higher the scores, the larger the percent increase of bottles sold for treated wines. The coefficient of interest is the one for the row "Score Level X Treated Store X Treated Period". We see that if the score displayed increases by one, then the number of bottles sold increases by 0.2 percent. Comparing columns (2) and (3) treatment effects, looking at those in the row "Treated Store X Treated Period," we see that demand for wines below a score of 81 (in the first score quartile) does not increase significantly (estimate of 0.005) while, for scores 81 and higher, demand increases significantly on average by 5.8 percent. In sum, we find a significant positive average consumer response to expert opinion labels for wine and significant demand increases for higher scoring wines.

5.3 Testing for price and wine quality belief in pre-period on treated wines

Table 4 presents the estimates of equation (1) for treated wines for four subsets of the data, namely for all four combinations of "high/low" scores-prices. The first column uses wines with scores in the higher quartiles (scores 81 or larger) and low price (that is, price in the lowest quartile); in column 2 we present estimates using wines in the high score and high price set; in column 3 we use data on wines in the low quality and low price set; and then column 4 uses data on wines in the low quality and high price set. The coefficient of interest is the difference in differences interaction coefficient, the β_4 in equation (1) which corresponds

[11] We have estimated alternative specifications using the score and price as continuous variables and interactions and higher order terms to allow for a non-linear relationship for the average treatment effects. We find that higher scores and higher prices are correlated with higher treatment effects, albeit not significantly, and while the point estimates do suggest a non-linear relationship for the treatment effect of price and score levels, the coefficients are not significant.

to the row "Treated Store X Treated Period." The null hypothesis is that demand for each subset of wines does not change significantly with our treatment. We only reject the null that there is no treatment effect for low price and high score wines, given that the coefficient on the "Treated Store X Treated Period" interaction, β_4, is positive and significant in column (1) but not statistically significant in the other columns.

5.4 Testing for price and wine region quality belief in pre-period on treated wines

We start with imported wines first in Table 5 and then turn to domestic wines in Table 6. In so doing we distinguish not only among the four possible score-price range combinations, but also dig deeper into which wine regions of imported and domestic wines are responsible for the average effects estimated in Table 4. The coefficient of interest is the difference in differences interaction coefficient, the β_4 in equation (1) which corresponds to the row "Treated Store X Treated Period". The null is that demand for each subset of imported and domestic wines does not change significantly with the treatment.

Starting with the imported set of wines, Table 5 presents estimates of equation (1) for imported treated wines with scores 81 or larger in the top panel, and for imported treated wines with scores below 81 in the bottom panel. Both panels allow differentiating per price range and imported wine regions: imported wines from Australia and New Zealand are in column (1), imported wines from Europe (Italy and France) are in column (2), other imported wines are in column (3), the lowest priced quartile imported wines are in column (4) and the imported wines in higher price quartiles are in column (5).

Starting with the top panel of Table 5, which focuses on the set of wines in the highest score quartiles, we reject the null that there is no treatment effect for all imported regions.

In particular, the point estimates point to a significant drop in quantity sold for all imported wines given the negative and significant point estimates in the row "Treatment Store X Treatment Period" for columns (1) to (3). Looking at column (4) we find no significant effects for high price quartile imported wines but estimate there to be a significant 73 percent drop in low priced and high score imported wines.

Turning to the bottom panel of Table 5 which focuses on the set of imported wines in the lowest score quartiles, we are only left with imported wines from Australia and New Zealand in the treated set. For these wines, we reject the null that there is no treatment effect for these two imported regions. In particular, the point estimates point to a significant increase in quantity sold for Australian and New Zealand imported wines, as shown by the positive and significant point estimates in the row "Treatment Store X Treatment Period" for column (1). Looking at columns (4) and (5), we find no significant effects when distinguishing among high price quartile and low price imported wines.

Turning now to the domestic set of wines, Table 6 presents estimates of equation (1) for domestic treated wines, which are all California wines. The top panel has California wines with scores 81 or larger; California treated wines with scores below 81 are in the bottom panel.. Both panels allow differentiating per price range and California wine regions: Napa and Sonoma regions are in column (1), other California regions are in column (2), the lowest priced quartile California wines are in column (3) and the California wines in higher price quartiles are in column (4). The coefficient of interest is, as before, the difference in differences interaction coefficient, which corresponds to the row "Treated Store X Treated Period". The null is that demand for each subset of wines in each column does not change significantly with the treatment.

Starting with the top panel of Table 6 which focuses on the set of wines in the highest score quartiles, we cannot reject the null of no treatment effect for the Napa and Sonoma regions but we reject the null that there is no treatment effect for other California regions. In particular, the point estimates in column (2) show a significant increase in quantity sold for other California region wines of 5.4 percent. Looking at column (4), we find significant effects for low price quartile and high scoring California wines, and estimate there to be a significant 18.1 percent increase in low priced and high score California wines.

Turning to the bottom panel of Table 6, which focuses on the set of California wines in the lowest score quartiles, we are only left with regions in California other than Napa and Sonoma in the treated set. For these wines, we cannot reject the null that there is no treatment effect for these other California regions. Looking at columns (3) and (4), we find no significant effects when distinguishing among high price quartile and low price California wines.

In sum, for the high score and low price subset of domestic wines (driven by regions other than Napa and Sonoma), demand increases by 18.1 percent due to the treatment. This means that we can reject that low price signals high quality for those wines. Predictions by Mahenc (2004), Spence (1976), and Tellis and Wernerfelt (1987) do not hold. Moreover, for the high score high price subset of domestic wines, demand does not increase significantly, evidence consistent with Shapiro (1983) and Bagwell and Riordan (1991) that high prices signal high quality for California wines. [12] This is reinforced by the the insignificance of the

[12] Producers of high quality products might not be profit-maximiser but instead utility-maximisers. As Morton and Podolny (2002) show, this implies lower price for products sold by utility maximisers. But this is true before and after the treatment, as the experiment does not affect the pricing strategy of wine products. thus, this does not affect our results. We empirically show that, in the context of a market where producers could

increase in demand for low score low price quartile California wines. In contrast, the demand for imported wines drops significantly for all regions due to the treatment for high scoring imported wines, suggesting that consumers switch away from high scoring foreign wines to domestic wines with high scores. This is consistent with consumers associating domestic wines with lower quality than imported wines – a belief that changes once the scores are revealed, leading demand for domestic wines to increase and demand for imported wines to drop.

5.5 *Belief updating due to spillover effects within and across treatment status*

Here, we want to investigate whether, when scores are revealed under the unchanged (pre and post experiment) price distribution of available choices, brand and variety, consumers update the way they use price in conjunction with all other attributes in order to infer quality in the post period. We test for belief updating in the form of spillover effects on unlabeled wines and distinguish by price quartile.

We compute a variable that features the average score of treated wines in the same brand of each wine, and a variable that features the average score of treated wines in the same variet. This is important because the wine bottles are organized on the shelves by varieties rather than by wine brand. The test for spillovers is to test whether the magnitude of treatment within brand causes demand to increase or not for wines in the same brand, and to perform the same empirical analysis within variety, taking advantage of the variation in the intensity of treatment by brand and variety and also of the variation in average scores for treated wines by brand and variety.

have heterogenous objective functions, the price-quality expectation relationship remains.

There are a total of 375 wine brands, where 297 of them have none of their products treated. However, for the remaining 76 wine brands, we have 6.37 percent of treated wines within a brand on average, with a standard deviation of 15.6. There are four brands that have all their available wines treated, and the remaining 72 brands have intensity of treatments ranging between 4.5% and 66%. The intensity of treatment also varies among the 14 different varieties. On average, 6.99% of wines in a variety are treated. Four varieties have no wines treated while, for the remaining ten varieties, intensity of treatment ranges between 4.76% and 22% of wines treated. In terms of average scores, among the 10 varieties that are treated, the average score ranges between a variety with an average score of 81.7 and a variety with an average score of 89. Average brand level scores range from 78 to 89 for their treated products, and the brand level average score is 83.6 with a standard deviation of 2.65.

Table 7 investigates within brand spillovers and reports the estimates in four columns according to treatment status and score levels for each subgroup of wines in the sample. Column (1) has treated wines with scores lower than 81, column (2) has treated wines with scores 81 or higher, column (3) has untreated wines with scores 81 or higher and column (4) has the group of wines without any scores. The last two columns (5) and (6) break up the spillover effects into untreated wines by price quartile.

The coefficients of interest are the ones associated with “Treatment X Average Score by Brand” and also “Treatment X Percent Treated by Brand” in each column. Given that the coefficients are not significantly different from zero in column (1) for the average and intensity of treatment by brand rows, we reject within-brand spillovers for treated wines of low score level. However, for the high scoring treated wines, we find that there are significant positive spillovers. In particular, we find that a one percent increase in the intensity of treated wines in a brand portfolio significantly causes the number of wine bottles to increase, with a significant point estimate of 96.9%, as shown in column (2). Given that among the treated brands, the range of treatment by brand is 4.5% and 66%, if the percentage of wines treated by brand increased to 66%, wine bottles would increase by 63.9% for each wine in the same brand that was treated.

For the untreated group that has scores higher than 81, in column (3), we reject the null of no spillovers. The coefficient of average score of treated wines in the same brand is positive and significant, meaning that the high quality spills over to untreated wines belonging to the same brand portfolio of high average scored brands. Moreover, the coefficient associated with the percent treated by brand is negative and significant. This implies that consumers perceive the wines that receive no labels as being of low quality. While brands that receive high average scores benefit from positive spillovers to the brand's score but untreated wines, the more wines that receive a score within a brand, the more consumers perceive the wines receiving no labels as being of poor quality.

For the un-scored wines in column (4), we reject the null of no spillovers. In fact, the estimates of average score and percent treated within brand spillover measures are both significant and positive. This is consistent with consumers inferring that, if a product belongs to a brand that in general has high scores or is highly labeled, then the unlabeled product must be good. As a result, they associate a brand that includes high-scoring products with good quality for the brand. Both columns (3) and (4) measure spillovers into sales of wines that do not receive a treatment label. While we see that having a product in its brand that received a higher average score has positive and significant spillovers for both untreated products with scores (in column 3) and untreated products without scores (in column 4), it is quite interesting that we see differences between column (3) and column (4) in terms of spillovers by intensity of treatment. For the group that never received a score (group in column 4) we find that they benefit from high intensity of treatment within its brand portfolio, while in column (3) it is the exact opposite, given the negative and significant point estimates associated with the intensity of treatment within brand. These are different types of wines: the sample of wines in column (3) has scores higher than 81 and the sample in column (4) would likely get a score less than 81 if scored. As a take away it appears that the spillover effects from the brand level intensity of treatment significantly benefit the lower-quality untreated wines relative to the higher scoring wines.

Lastly we investigate whether there are differential spillovers depending on wine prices. In column (5), the sample consists of untreated wines in the lowest price quartile, while column (6) has wines in higher price quartiles. We cannot reject the null of no spillover effects for both samples given the non-significant coefficients in both columns (5) and (6) of Table 7 for rows “Treatment X Average Score by Brand” and “Treatment X Percent Treated by Brand.”

Table 8 presents the estimates used to test for spillovers within varieties. Table 8 is organized in the same way as Table 7 and only the coefficients in column (2) reject a conclusion that there are no spillovers. For the sample of wines with scores 81 or higher, we find that higher average scores within variety reduce demand for treated wines in the same variety. Additionally, the higher the intensity of treatment within variety, the larger the increase in demand for treated wines of scores 81 or higher.

In columns (5) and (6) of Table 8, we investigate whether there are differential spillovers depending on wine prices. In column (5), the sample consists of untreated wines in the lowest price quartile, while column (6) has wines in higher price quartiles. We cannot reject the null of no spillover effects for higher price ranges in column (6) but we reject the null for lower price ranges in column (5). Specifically, we find evidence consistent with larger brand spillover effects in the lower price quartile given the positive and significant point estimate in column (5) and row "Treatment X Average Score by Variety" of 0.011, which is consistent with the findings in Costanigro et al, (2010). Moreover, the greater the intensity of the treatment for a variety, the more negative the effect on demand for wines in low price ranges, given the negative and significant point estimate in column (5) in row "Treatment X Percent Treated by Variety" of -103.11.

5.6 Additional specifications and robustness

We also investigate overall treatment effects resulting from this experiment on store level Total Quantity and Total Revenue. The estimates in Table A.1 in the Appendix show that there are no significant total effects in terms of changes in quantity, in column (4), or revenue, in column (5), due to the treatment, nor are there significant effects when we estimate the equation (1) by treatment status separately, as can be seen in columns (1) to (3).

Given that the disclosure of scores have to be seen within region and within variety, in alternative specifications we define high and low quality in the context of each specific variety. Results are robust and consistent with our main findings when defining high and low quality in terms of score quartiles among all wines.[13].

[13]Results and Tables are presented in the online appendix. First we compute quartiles by variety for treated wines and note that there is little variation for the set of wines in the data. When breaking up the score ranges for the set of wines we treat, we happen to have a similar range of displayed scores across all varieties, as can be seen in the appendix.

6 Conclusion

We empirically estimate the effect on wine demand of revealing expert opinion information about the quality of a subset of products in the choice set of consumers. We not only estimate average effects, but we break up the effects by groups of wines without and without expert information, to investigate and test possible theoretical mechanisms as to how expert opinion labels can reveal the way consumers make inferences about quality and compare products in the context of incomplete information.

Our estimates suggest there is a positive and significant overall average effect and that demand increases more for higher score wines than for lower score wines (as in Hilger et al., 2011). Our contribution lies in new results that suggest that higher scores matter more for prices in the lower quartile of the overall wine price distributions using revealed choices. Our analysis rules out that consumers do not perceive low price as signaling high quality. Our findings are instead consistent with consumer behavior where demand does not move for higher price quartiles once quality is revealed as consumers infer high quality for high prices. These empirical findings allows the testing of theoretical predictions. Our evidence is consistent with high price - high quality expectations as in Bagwell and Riordan (1991) and Shapiro (1983). This paper contributes to the empirical literature as it uses actual revealed preferences where price is perceived as being positively correlated with wine quality when consumers have incomplete information. We find that consumers update their beliefs of low prices signaling lower quality, all else equal, given that they increase demand for lower quartile priced wines that have a high score. This result is consistent with previous literature findings of wine tasting experiments and choice surveys (see e.g., Mastrobuoni, et al, 2014).

We also find positive spillover effects of this experimental treatment within brand for untreated wines, especially if the average score of wines within brand is high. However, if there is a higher percentage of wines within a given brand that are treated - that is, the brand is intensively treated - we find that there are significant negative spillovers for untreated wines. We do not find significant negative spillovers within varieties.

In terms of beliefs about quality related to wine region of production, we find that consumers update their beliefs about California wines not originating from the Napa and Sonoma regions, if they receive scores of 81 or higher. Moreover, consumers switch away from high scoring foreign wines to California wines with high scores originating from regions other than Napa or Sonoma. This is consistent with consumers associating domestic wines from these regions with lower quality than imported wines – a belief that changes once the scores are revealed, leading demand for domestic wines to increase and demand for imported wines to drop.

As a limit, given that we do not have access to consumer panel data we cannot investigate whether some subset of consumers were previous consumers of wine and whether they have knowledge of the wines they buy other than the labels provided by the seller. We leave this as an avenue for future research in a retail setting where such data exist.

7 References

Ackerberg, D. (2003). Advertising, Learning, and Consumer Choice in Experience Good Markets: An Empirical Examination. *International Economic Review* 44(3): 1007-1040.

Akerlof, G. A. (1970). The Market for 'Lemons': Quality Uncertainty and the Market

Mechanism. *The Quarterly Journal of Economics* 84(3): 488-500.

Ali, H. H. and Nauges, C. (2007). The pricing of Experience Goods: The example of En Primeur Wine. *American Journal of Agricultural Economics* 89(1): 91-103.

Bagwell, K. and Riordan, M.H. (1991). High and Declining Prices Signal Product Quality. *American Economic Review* 81: 224-239.

Bertrand M., Duflo, E. and Mullainatha, S. (2004). How much should we trust difference-in-differences estimates? *Quarterly Journal of Economics* 119(1): 249-275.

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

Chocarro, R. and Cortiñas, M. (2013). The impact of expert opinion in consumer perception of wines. *International Journal of Wine Business Research* 25(3): 227-248.

Costanigro, M., McCluskey, J. J. and Goemans, C. (2010). The economics of nested names: name specificity, reputations, and price premia. *American Journal of Agricultural Economics* 92(5): 1339-1350.

Crooks, E. (1989). Alcohol consumption and taxation. *The Institute for Fiscal Studies*, London, UK.

Davis T.R., Ahmadi-Esfahani, F. Z. and Iranzo, S. (2008). Demand under product differentiation: an empirical analysis of the US wine market. *The Australian Journal of Agricultural and Resource Economics* 52: 401-417

Dubois, P. and Nauges, C. (2010). Identifying the Effect of Unobserved Quality and Experts' Reviews in the Pricing of Experience Goods: Empirical Application on Bordeaux Wine. *International Journal of Industrial Organization* 28(3): 205-212.

Duffy, M. (2002). On the estimation of an advertising-augmented, cointegrating demand system. *Economic Modelling* 20(1): 181-206.

Friberg R. and Grnqvist, E. (2012). Do Expert Reviews Affect the Demand for Wine? *American Economic Journal: Applied Economics* 4(1): 193-211.

Gneezy, A., Gneezy, U. and Lauga, D.O. (2014). A reference-dependent model of the pricequality heuristic. *Journal of Marketing Research* 51(2): 153-164.

Grankvist, G. and Biel, A. (2007). Predictors of purchase of eco-labelled food products: A panel study. *Food Quality and Preference* 18: 701-708.

Grossman, S.J. (1981). The Informational Role of Warranties and Private Disclosure about Product Quality. *Journal of Law and Economics* 24: 461-89.

Hadj Ali, H., Lecocq, S. and Visser, M. (2010). The Impact of Gurus: Parker Grades and en primeur Wine Prices, *Journal of Wine Economics* 5(1): 22-39.

Hensher D., Louviere, J. and Swait, J. (1998). Combining sources of preference data. *Journal of Econometrics* 89(1-2): 197-221.

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

Jin, G. Z. and Leslie. P. (2003). The Effects of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards. *Quarterly Journal of Economics* 118: 409-451.

Jover, A. J. V., Llorns Montes, F. J. and Fuentes, M. F. (2004). Measuring perceptions of quality in food products: the case of red wine. *Food Quality and Preference* 15: 453-469.

Kiesel, K., and Villas Boas, S. B. (2007). Got Organic Milk? Consumer Valuations of Milk Labels after the Implementation of the USDA Organic Seal. *Journal of Agricultural &*

Food Industrial Organization 5(1): 1-38.

Mahenc, P. (2004). Influence of informed buyers in markets susceptible to the lemons problem. *American Journal of Agricultural Economics* 86(3):649-659.

Mastrobuoni, G., Peracchi, F. and Tetenov, A. (2014). Price as a signal of product quality: Some experimental evidence. *Journal of Wine Economics* 9: 135-152.

Meyer, B. D. (1995). Natural and Quasi-Experiments in Economics. *Journal of Business & Economic Statistics* 13(2): 151-161.

Montgomery, C.A. and Wernerfelt, B. (1992). Risk reduction and umbrella branding. *Journal of Business* 1: 31-50.

Morton S. F. and Podolny, J. (2002). Love or Money? The Effects of Owner Motivation in the California Wine Industry. *The Journal of Industrial Economics* 50(4): 431-456.

Ozretic-Dosen, D., Skare, V. and Krupka, Z. (2007). Assessments of country of origin and brand cues in evaluating a Croatian, western and eastern European food product. *Journal of Business Research* 60: 130-136.

Palma, D., Ortuzar, J. D., Casaubon, G., Rizzi, L.I. and Agosin, E. (2013). Measuring Consumer Preferences Using Hybrid Discrete Choice Models. *AAWE WORKING PAPER* 137.

Reinstein, D. A. and Snyder, C. M. (2005), The influence of expert reviews on consumer demand for experience goods: a case study of movie critics. *The Journal of Industrial Economics* 53: 2751.

Schnettler, B., Vidal, R., Silva, R., Vallejos, L. and Seplveda, L. (2009). Consumer willingness to pay for beef meat in a developing country: The effect of information regarding country of origin, price and animal handling prior to slaughter. *Food Quality and Preference*

20: 156-165.

Scozzafava, G., Boncinelli, F., Contini, C., Romano, C., Gerini, F. and Casini, L. (2016). Typical vine or international taste: wine consumers dilemma between beliefs and preferences. *Recent patents on food, nutrition and agriculture* 8(1): 31-38.

Shapiro, C., (1983). Premiums for high quality products as returns to reputations. *The Quarterly Journal of Economics* 98(4):659-79.

Sorensen, A. T. and Rasmussen, S.J. (2004). Is Any Publicity Good Publicity? A Note on the Impact of Book Reviews. *Working paper, Stanford University* 1-16.

Spence, M. (1976). Informational aspects of market structure: An introduction. *The Quarterly Journal of Economics* 90(4): 591-597.

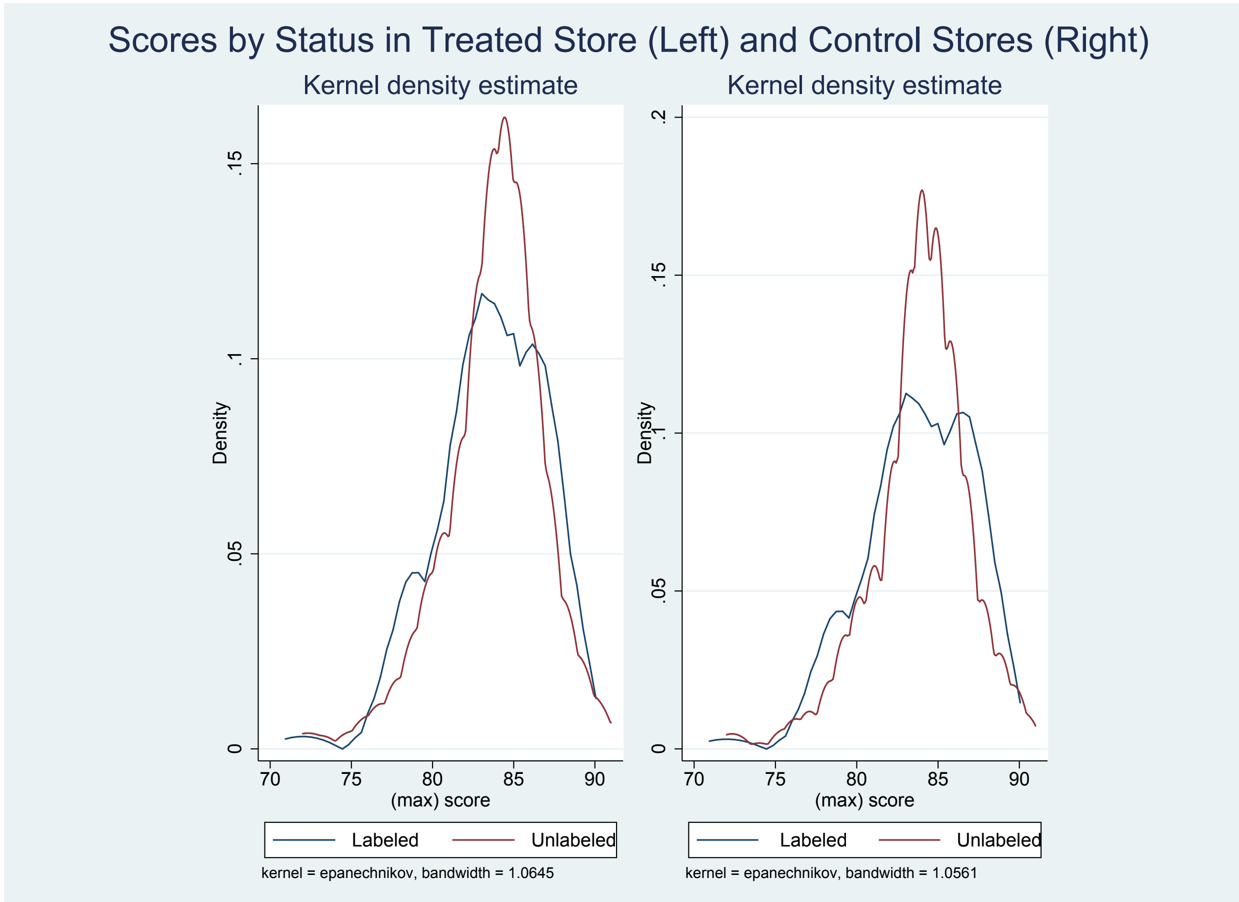
Tellis, G. J. and Wernerfelt, B. (1987). Competitive price and quality under asymmetric information. *Marketing Science* 6(3): 240-253.

Veale, R. (2008). Sensing or knowing? Investigating the influence of knowledge and self-confidence on consumer beliefs regarding the effect of extrinsic cues on wine quality. *International Journal of Wine Business Research* 20(4): 352-366.

Veale, R. and Quester, P. (2009). Do consumer expectations match experience? Predicting the influence of price and country of origin on perceptions of product quality. *International Business Review* 18: 134-144.

Whitehead, J. C., Pattanayak, S. K., Van Houtven, G. L. and Gelso, B. R. (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: 872-908.

Figure 1: Histogram of Scores by Status of Wines in Treated and Control Stores



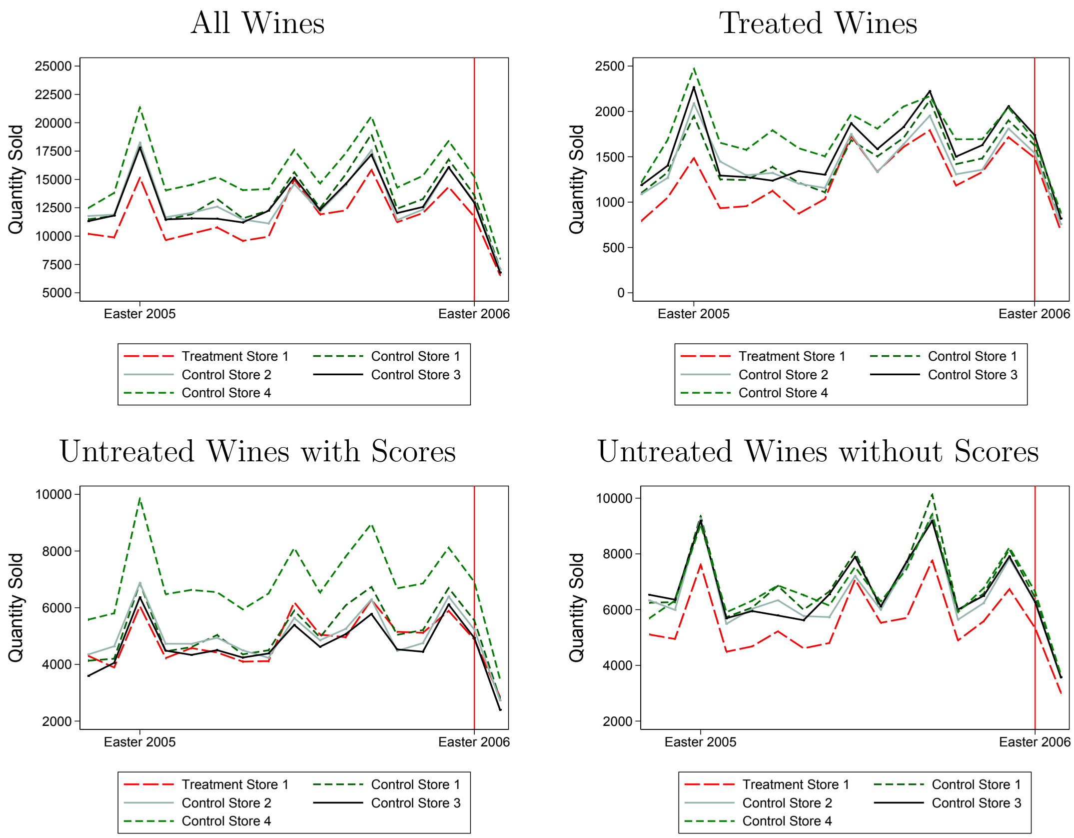
Note: There are scores available for wines pertaining to two types of status: Labeled status means these wine products are labeled and treated in the treated store. Unlabeled Status are the wines that have scores but do not receive a label with a score, thus are not treated in the Treated Store. The left panel displays the kernel density estimates of the score distribution for labeled and unlabeled wine products in the treated store. The right panel displays the kernel density estimates of the score distribution in control stores for the same group of wine products labeled and unlabeled (receiving a label and unlabeled in the treated store), given that we can see the same products in the control stores, which will be our counterfactuals in the difference in difference estimation strategy. Kolmogorov-Smirnov (KS) test for equality of the distribution in each of the panels cannot reject the equality of scores distribution in the treated and also in the control stores between the labeled and unlabeled status of wines , given that KS (pvalue) is for the Treated Store= 0.0987 (0.481) and for the Control Stores =0.0935 (0.518).

Table 1: Summary Statistics of Data for Wines by Treatment Status for Treated and Control Stores

	(1)	(2)	(3)	(4)	(5)	(6)
		Treated Store			Control Stores	
	Set 1 Wines Treated	Set 2 Wines Untreated With Scores	Set 3 Wines Without Scores	Set 1 Wines Treated	Set 2 Wines Untreated With Scores	Set 3 Wines Without Scores
Quantity (March)	16.99	25.60	11.97	19.19	29.06	13.42
	(26.01)	(44.90)	(21.57)	(26.01)	(53.28)	(23.75)
Quantity (April)	14.88	22.08	10.31	16.93	24.56	11.21
	(22.23)	(40.35)	(17.96)	(22.53)	(47.45)	(19.70)
Price (March)	10.98	10.15	10.98	11.23	10.38	11.36
	(5.00)	(4.81)	(6.43)	(5.02)	(4.91)	(6.50)
Price (April)	10.96	10.07	10.93	10.97	10.31	11.38
	(5.15)	(4.73)	(6.14)	(4.83)	(5.13)	(6.40)
% discounted (March)	0.91	0.99	0.75	0.88	0.99	0.74
% discounted (April)	0.88	0.99	0.78	0.89	0.98	0.75
Score	83.13	83.65		83.05	83.68	
	(3.20)	(3.20)		(3.29)	(3.01)	
% red	0.58	0.45	0.50	0.59	0.46	0.51
% white	0.35	0.48	0.40	0.34	0.46	0.39
Number Wines	101	230	563	102	235	599
Number Observations	2562	5721	16194	11058	24578	72871
		By Brand			By Variety	
Average Score		83.05			83.59	
Min - Max Score		78 - 89			81.7 - 89	
Average % Treated		6.37			6.99	
Min - Max % Treated		0-100			0 - 22	
		Number Brands 375			Number Varietals 14	

Sample descriptive statistics are reported. First three columns for Treated Store, next three for Control stores.
Set 1 of wines= products that are labeled in treated store. Set 2 products that have scores but are not labeled in treated store. Set 3 products without scores.
The value reported for quantity, price and score are the sample average. Standard deviation are in parenthesis.

Figure 2: Monthly Quantities Sold by Store



Note: Treatment period is the Vertical Red Line.

Table 2: Difference-in-Difference: Average Effect of Expert Opinion Labels on Retail Wine Sales

	(1) Quantity Sold (Q)	(2) Q	(3) ln(Q)
Treated Store X Treated Period	1.520***	1.507***	0.052
	(0.245)	(0.240)	(0.046)
Treated Store	-3.104***	-3.419***	-0.186***
	(0.261)	(0.266)	(0.015)
Treated Period	-1.851	-3.277	-0.193
	(3.345)	(2.797)	(0.171)
Price		-2.756***	
		(0.127)	
Discount Dummy		2.717***	0.557***
		(0.592)	(0.048)
ln(Price)			-2.671***
			(0.135)
Mean of Dep. Variable	15.118	15.118	1.943
Num of Obs.	13619	13619	13617
R squared	0.462	0.498	0.595
Product FE	X	X	X

This Table presents the estimates of equation (1) for treated wines.
The coefficient of interest is the difference in differences interaction coefficient, the β_4 in equation (1), which corresponds to the row "Treated Store X Treated Period". The null is that demand for wine does not change significantly with the treatment.
Clustered errors in parentheses at the month level. Controls are best 4 stores. The dependent variable is the quantity by product sold per month except the last column, which is ln quantity sold.
$*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table 3: Difference-in-Difference: Score Level Effect of Expert Opinion Labels on Retail Wine Sales

	(1) ln(Q)	(2) ln(Q) and Score < 81	(3) ln(Q) and Score > 80
ln(Price)	-2.671***	-2.976***	-2.590***
	(0.135)	(0.217)	(0.150)
Discount Dummy	0.556***	0.481***	0.572***
	(0.048)	(0.051)	(0.055)
Score Level X Treated Store X Treated Period	0.002***		
	(0.000)		
Treated Store X Treated Period	-0.060	0.005	0.058**
	(0.060)	(0.156)	(0.025)
Treated Store	-0.186***	-0.248***	-0.172***
	(0.015)	(0.029)	(0.018)
Treated Period	-0.193	-0.317*	-0.165
	(0.171)	(0.163)	(0.171)
Mean of Dep. Variable	1.943	2.154	1.893
Num of Obs.	13617	2609	11008
R squared	0.595	0.553	0.600
Product FE	X	X	X

This Table presents the estimates of equation (1) for treated wines and also interacts the Treatment with the score Level in Column (1).
The coefficient of interest is the difference in differences interaction coefficient interacted with score, which corresponds to the row "Score Level X Treated Store X Treated Period". The null is that demand does not change significantly with the treatment when score increases.
Rather then assuming the linear score treatment interaction in column (1), we break the sample of wines in two and re-estimate equation (1).
In Column (2) we use only wines with Scores Less than 81, that is in the first score quartile, and Column (3) we use Wines with Scores greater than or equal to 81.
The dependent variable is the log of quantity in all the columns.
Clustered errors in parentheses are at the month level.
$*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table 4: Heterogeneous Score Level Effect of Expert Opinion Labels on Retail Wine Sales Scores 81 or Higher

	(1) High Quality Low Price	(2) High Quality High Price	(3) Low Quality Low Price	(4) Low Quality High
Ln (Price)	-2.873***	-2.554***	-3.287***	-2.571***
	(0.403)	(0.137)	(0.208)	(0.404)
Discount Dummy	0.667***	0.561***	0.449***	0.545***
	(0.094)	(0.056)	(0.092)	(0.081)
Treated Store X Treated Period	0.181***	0.039	-0.050	0.093
	(0.047)	(0.025)	(0.253)	(0.088)
Treated Store	-0.231***	-0.162***	-0.287***	-0.206***
	(0.045)	(0.017)	(0.050)	(0.057)
Treated Period	-0.231	-0.156	-0.563***	-0.086
	(0.157)	(0.173)	(0.148)	(0.191)
Mean of Dep. Variable	2.387	1.813	2.579	1.765
Num of Obs.	1545	9463	1245	1364
R squared	0.567	0.595	0.553	0.427
Product FE	X	X	X	X

This Table presents the estimates of equation (1) for treated wines for four subsets of the data, namely for all combinations of "high/low" scores-prices. The coefficient of interest is the difference in differences interaction coefficient, the β_4 in equation (1), which corresponds to the row "Treated Store X Treated Period". The null is that demand for each subset of wines does not change significantly with the treatment.

The dependent variable is the Log quantity in all columns. We define Low Quality as a score in the first score quartile (below 81) and high quality for higher quart

We present the results for high quality wines in Column (1) and (2) and for low quality wines in columns (3) and (4).

We define Low Price as a price in the first quartile, and high price in the higher quartiles.

Clustered errors in parentheses are at the month level.

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

Table 5: Heterogeneous Score Level Effect of Expert Opinion Labels on Imported Retail Wine Sales by Imported Region

	(1) Imported NZ+Australia	(2) Imported Europe	(3) Imported Other	(4) First Price Quartile	(5) Highest Price Quartiles
	Wines in High Score Quartiles (Scores 81 or Higher)				
Ln(Price)	-2.717***	-3.763***	-0.336	-2.956***	-2.782***
	(0.507)	(0.472)	(0.720)	(0.782)	(0.236)
Discount Dummy	0.537***	0.251**	0.719***	1.029***	0.353***
	(0.141)	(0.093)	(0.179)	(0.258)	(0.057)
Treated Store X Treated Period	-0.237***	-0.281***	-0.511***	0.110	-0.732***
	(0.072)	(0.071)	(0.080)	(0.107)	(0.107)
Treated Store	-0.101*	-0.189***	-0.055	-0.099	-0.054
	(0.051)	(0.055)	(0.097)	(0.090)	(0.096)
Treated Period	-0.173	-0.004	-0.572***	-0.168	-0.718***
	(0.133)	(0.208)	(0.109)	(0.192)	(0.117)
Mean of Dep. Variable	1.937	1.474	0.872	2.676	1.319
Num of Obs.	1290	634	199	595	1528
R squared	0.606	0.600	0.217	0.505	0.510
	Wines in Lowest Score Quartile (Scores Less than 81)				
Ln (Price)	-1.423**			-3.287***	-2.571***
	(0.646)			(0.208)	(0.404)
Discount Dummy	0.750**			0.449***	0.545***
	(0.309)			(0.092)	(0.081)
Treated Store X Treated Period	1.068***			-0.050	0.093
	(0.137)			(0.253)	(0.088)
Treated Store	-0.363**			-0.287***	-0.206***
	(0.159)			(0.050)	(0.057)
Treated Period	-1.034***			-0.563***	-0.086
	(0.263)			(0.148)	(0.191)
Mean of Dep. Variable	1.690			2.579	1.765
Num of Obs.	151			1245	1364
R squared	0.218			0.553	0.427

This Table presents the estimates of equation (1) for imported treated wines by country of origin. The coefficient of interest is the difference in differences interaction coefficient, the β_4 in equation (1) which corresponds to the row "Treated Store X Treated Period". The null is that demand for wine does not change significantly with the treatment.

The dependent variable is the Log quantity in all columns. All specifications have product fixed effects. Clustered errors in parentheses at the month level. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table 6: Heterogeneous Score Level Effect of Expert Opinion Labels on Domestic (California) Retail Wine Sales

	(1) Napa and Sonoma	(2) Other CA Regions	(3) First Price Quartile	(4) Highest Price Quartiles
	Wines in High Score Quartiles (Score 81 or Higher)			
Ln (Price)	-2.590***	-2.589***	-2.873***	-2.554***
	(0.536)	(0.151)	(0.403)	(0.137)
Discount Dummy	0.722***	0.566***	0.667***	0.561***
	(0.115)	(0.055)	(0.094)	(0.056)
Treated Store X Treated Period	0.144	0.054**	0.181***	0.033
	(0.120)	(0.021)	(0.047)	(0.020)
Treated Store	-0.223**	-0.169***	-0.231***	-0.158***
	(0.076)	(0.017)	(0.045)	(0.016)
Treated Period	-0.208	-0.163	-0.231	-0.153
	(0.138)	(0.173)	(0.157)	(0.176)
Mean of Dep. Variable	1.555	1.912	2.387	1.813
Num of Obs.	582	10426	1545	9463
R squared	0.493	0.603	0.567	0.595
	Wines in Lowest Score Quartile (Score Less than 81)			
Ln(Price)		-2.976***	-3.287***	-2.571***
		(0.217)	(0.208)	(0.404)
Discount Dummy		0.481***	0.449***	0.545***
		(0.051)	(0.092)	(0.081)
Treated Store X Treated Period		0.005	-0.050	0.093
		(0.156)	(0.253)	(0.088)
Treated Store		-0.248***	-0.287***	-0.206***
		(0.029)	(0.050)	(0.057)
Treated Period		-0.317*	-0.563***	-0.086
		(0.163)	(0.148)	(0.191)
Mean of Dep. Variable		2.154	2.579	1.765
Num of Obs.		2609	1245	1364
R squared		0.553	0.553	0.427

This Table presents the estimates of equation (1) for domestic treated wines by region. The coefficient of interest is the difference in differences interaction coefficient, the β_4 in equation (1) which corresponds to the row "Treated Store X Treated Period". The null is that demand for wine does not change significantly with the treatment. All domestic wines in the treated sample are CA wines. In Column (1) we have Napa and Sonoma Wines, in column (2) wines from other CA regions, column (3) features estimates for Lowest Quartile Priced Wines, and column (4) higher Quartile Price Wines. The dependent variable is the Log quantity in all columns. All specifications have product fixed effects. Clustered errors in parentheses at the month level. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table 7: Spillover Effects of Expert Opinion Labels on Retail Wine Sales Within Brand

	(1)	(2)	(3)	(4)	(5)	(6)
Ln (Price)	-2.66***	-2.98***	-0.92***	-2.06***	-2.79***	-2.61***
	(0.17)	(0.22)	(0.27)	(0.08)	(0.27)	(0.17)
Discount Dummy	0.59***	0.48***	0.44***	0.45***	0.69***	0.53***
	(0.06)	(0.05)	(0.07)	(0.04)	(0.11)	(0.04)
Treatment	0.17	-2.59	0.19	-0.29***	0.06	-0.06
	(0.55)	(1.83)	(0.35)	(0.08)	(0.22)	(0.11)
Treatment X Avg Score by Brand	-0.00	0.03	0.02***	0.003**	-0.00	-0.00
	(0.01)	(0.02)	(0.01)	(0.00)	(0.00)	(0.00)
Treatment X Percent Treated by Brand	-4.76	96.94**	-2198.94**	25.53***	89.53	37.29
	(4.59)	(39.63)	(962.63)	(5.33)	(82.73)	(28.19)
Mean of Dep. Variable	1.95	2.15	1.04	1.72	2.51	2.21
Num of Obs.	9550	2609	701	24270	4731	11379
R squared	0.61	0.56	0.28	0.604	0.58	0.71
Product FE	X	X	X	X	X	X

This Table presents the estimates of equation (1) for different sets of wines and interacts "Treatment" with Average Score by Brand and Percent Treated by Brand, respectively.
The coefficients of interest are the difference in differences coefficient of the interaction of Treatment with those variables, which corresponds to the rows "Treatment X Average Score by Brand" and "Treatment X Percent Treated by Brand." The null is that there are no spillovers, in that demand does not change significantly with the treatment the more a brand is treated and the higher the average score within brand.
The dependent variable is the Log quantity in all columns.
Column (1) has treated wines (product set 1) with scores lower than 81,
Column (2) has treated wines (product set 1) with scores 81 and higher,
Column (3) has untreated wines with scores (product set 2) 81 and higher,
Column (4) has untreated wines without scores (product set 3), for which we do not have available score information from experts.
Column (5) has untreated wines with scores (product set 2) with low prices
Column (6) has untreated wines with scores (product set 2) with high prices.
All lower order interactions-combinations were in the regression but are omitted from this Table due to space.
Clustered errors in parentheses at the month level.
$*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table 8: Spillover Effects of Expert Opinion Labels on Retail Wine Sales Within Variety

	(1)	(2)	(3)	(4)	(5)	(6)
ln (Price)	-2.60***	-2.97***	-1.27***	-2.08***	-2.98***	-2.99***
	(0.15)	(0.22)	(0.16)	(0.06)	(0.23)	(0.13)
Discount Dummy	0.57***	0.48***	0.35***	0.42***	0.47***	0.48***
	(0.06)	(0.05)	(0.04)	(0.03)	(0.06)	(0.04)
Treatment	0.49	2.42***	4.67	0.11	-0.74***	0.25
	(0.49)	(0.48)	(3.07)	(0.15)	(0.24)	(0.42)
Treatment X Average Score by Variety	-0.01	-0.04***	-0.07	-0.00	0.01***	-0.00
	(0.01)	(0.01)	(0.06)	(0.00)	(0.00)	(0.01)
Treatment X Percent Treated by Variety	-96.59	811.14*	-171.50	-44.12	-103.11**	-2.40
	(70.40)	(376.05)	(667.44)	(79.49)	(37.52)	(35.87)
Mean of Dep. Variable	1.89	2.15	0.98	1.59	2.38	2.03
Num of Obs.	11008	2609	943	90588	8902	21211
R squared	0.60	0.55	0.28	0.56	0.56	0.68
Product FE	X	X	X	X	X	X

This Table presents the estimates of equation (1) for different sets of wines and interacts "Treatment" with Average Score by Variety and Percent Treated by Variety, respectively.
The coefficients of interest are the difference in differences coefficient of the interaction of Treatment with those variables, which corresponds to the rows "Treatment X Average Score by Variety" and "Treatment X Percent Treated by Variety." The null is that there are no spillovers, in that demand does not change significantly with the treatment the more a variety is treated and the higher the average score within variety.
The dependant variable is the Log quantity in all columns. Column (1) has treated wines (product set 1) with scores lower than 81,
Column (2) has treated wines (product set 1) with scores 81 and higher,
Column (3) has untreated wines with scores (product set 2) 81 and higher,
Column (4) has untreated wines without scores (product set 3), for which we do not have available score information from experts.
Column (5) has untreated wines with scores (product set 2) with low prices
Column (6) has untreated wines with scores (product set 2) with high prices.
All lower order interactions-combinations were in the regression but are omitted from this Table due to space.
Clustered errors in parentheses at the month level. The dependent
$*p < 0.10$, $**p < 0.05$, $***p < 0.01$