Journal of Wine Economics (2022), **17**, 177–208
doi:10.1017/jwe.2022.23

ARTICLE

Product differentiation and the relative importance of wine attributes: U.S. retail prices

Raj Chandra[1] and GianCarlo Moschini[2]

[1]Department of Economics, Iowa State University, Ames, IA 50011 and [2]Department of Economics and Center for Agricultural and Rural Development, Iowa State University, Ames, IA 50011
Corresponding author: GianCarlo Moschini, email: moschini@iastate.edu

Abstract
This paper investigates the relative importance of various attributes, including varietal, brands, and geographic origin, in explaining retail wine prices for the United States market. We use a metric based on the Shapely value, from cooperative game theory, in the context of an empirical hedonic price equation estimated using a large sample of retail wine sales for home consumption over the period 2007–2019. We find that brands alone explain more than 70% of the variation in wine prices, but geographic origin and varietals retain additional explanatory power. Furthermore, information about the geographic origin appears to be a considerably more important attribute than varietals.

Keywords: American viticulture areas; hedonic price functions; Shapley value; U.S. wine; wine prices
JEL Classifications: C20; L15; L66; Q11

ELWOOD: "… what kind of music do you usually have here?"
CLAIRE: "Oh, we got both kinds. We got country, and western."
— The Blues Brothers, 1980

I. Introduction

Product differentiation has emerged as a central theme in modern industrial organization. Strong market forces are behind the proliferation of product varieties. On the demand side, consumer heterogeneity plays a major role, with individual choices reflecting differences in underlying tastes and preferences as well as in consumers' income. On the supply side, product differentiation is a powerful tool for firms to soften the adverse profit consequences of price competition (Tirole, 1988). Much interest concerns the margins along which firms can profitably differentiate their products, with significant implications for business strategy, industry evolution, and public policy. Product differentiation has long been a distinctive feature of the wine industry. Although interest often centers on quality, along an implied vertical differentiation dimension, it is clear that horizontal differentiation catering to the heterogeneity of consumers is just as relevant. In this paper, we assess the determinants of product differentiation in the U.S. wine industry, with special emphasis on the role of geographic origin of wines.

Wine is a classic example of an experience good where consumers cannot ascertain the quality of the product before consumption (Storchmann, 2012), which presents a potentially deleterious structural asymmetric information problem (Akerlof, 1970). Several market mechanisms have emerged to deal with this situation; chief among them is firms' pursuit of strong brand identities that, by leveraging the notion of reputation, can foster repeat purchases (Shapiro, 1983). In addition to individual firms' branding strategies—which are supported by proprietary trademarks and are ubiquitous in most differentiated product markets—the wine industry has also developed a somewhat unique system of geographic origin labeling. This approach is based on France's historic experience with the "appellations" system (Meloni and Swinnen, 2013). The core premise is that major components of a wine's distinctiveness are the specific geo-climatic conditions of its production region, the historic grape varieties that characterize the region, and the traditional local practices of wine producers.

The evidence shows that, indeed, the system of appellations has been effective in ameliorating the market failures of asymmetric information in the French wine industry (e.g., Mèrel, Ortiz-Bobea, and Paroissien, 2021). In fact, geographic denominations of origin—essentially, collective labeling mechanisms that complement firms' own brands (Menapace and Moschini, 2012)—have been widely adopted in the European Union, which has institutionalized the associated notion of "terroir" into its food quality policy. The extent to which geographic origin labeling should matter for the rest of the world remains an open question, however, given that major elements of the European wine industry development are not common elsewhere. For example, varieties used to produce European wines are typically inherited as part of the local history (e.g., Sangiovese is used to produce Chianti wines), whereas, in the so-called New World, varieties are deliberately chosen by wine producers under fewer constraints.[1] In fact, the use of varietals has emerged as a key labeling and marketing tool for U.S., Australian, and South American wines.

A plethora of information is typically found on wine labels—often including brand, grape variety, geographic origin (possibly in the form of an appellation classification), alcohol content, and vintage. These intrinsic attributes, combined with other signals, such as price and expert rating, help consumers make wine choices and increase the likelihood of repeated purchases (Horowitz and Lockshin, 2002). Assessing the relative importance of such factors remains an active area of research, with implications that range from the design of food quality policies to the promotion and marketing strategies of individual firms. In particular, one can ask to what extent geography per se matters relative to brands (Schamel, 2006), or whether geography is redundant information once variety and brand information are known. The latter concept, in particular, has been articulated as a stylized difference between so-called Old World and New World wine industry strategies (Castriota, 2020).

[1]For example, the penetration of Bordeaux-style grapes in Tuscany—which contributed to the development of so-called Super-Tuscan wines—has been slow and constrained by appellation rules favoring traditional Chianti.

Somewhat in contrast with the accepted view that collective designation of origin labeling is more important for the Old World than the New World wine industry (Lockshin and Corsi, 2012), an own appellation system has taken roots in the United States and grown over time. In addition to the use of state or county labels to designate the origin of their wines, the United States has a federally recognized appellation of origin designation for its prestigious wine-growing regions known as American Viticulture Areas (AVAs). The first AVA was designated in 1980, on the heels of the 1976 "Judgement of Paris" that brought heightened recognition to California (and especially Napa Valley) wines on the global scene (Taber, 2005). The AVA program has grown significantly since then. As of March 2022, there are 261 approved AVAs in the United States. Whereas county and state labels follow the boundary of the named jurisdiction, AVAs are delimited by their distinctive geographic and climatic features, and can be in a single county (e.g., Napa Valley), can spread over more than one county (e.g., Los Carneros), and can cross state boundaries (e.g, Columbia Valley).

AVA, state, and/or county designations are now routinely found on U.S. wine labels, alongside many other informational elements. In fact, wine consumers are typically faced with a bewilderingly large, and expanding, choice set (Orth, Lockshin, and d'Hauteville, 2007). Hence, there is a longstanding interest in understanding the essential wine attributes that are chiefly associated with wine choice and their prices. In particular, given its relatively new history, the role of AVA designation in this context remains of considerable importance for the wine industry and policy makers.

In this paper, we investigate the relative importance of some main wine attributes in explaining wine prices in the U.S. market, with particular emphasis on the impact of AVA classifications. We use an extensive dataset from Nielsen's consumer panel data for the United States from 2007 to 2019. These household-based homescan data are founded on a nationally-representative panel of approximately 60,000 households per year, organized into 61 geographic areas, including 52 major markets and nine census divisions. The specific information we use pertains to wine purchases carried out by these households. For each observed purchase instance (shopping trip), we have data on the wine purchased at the very detailed Universal Product Code (UPC) level, as well as the price paid for the product. For each UPC, we extract information about some key attributes of the wine, including wine type, varietal, and geographic origin. The analysis we present focuses on U.S. wines sold in standard 750 ml bottles.

These data on actual prices paid and observed product characteristics are used to estimate hedonic price functions. The application of hedonic methods is common in the wine economics literature (Outreville and Le Fur, 2020). Although we do not advance the wine hedonic methodology per se, we believe our work exhibits some novel features. First, our results are based on a very large dataset broadly representative of the entire U.S. wine market. Much of the previous wine hedonics work focuses on quality, often limited to premium wines, and relies on limited datasets specifically suited to that purpose. We take a broader view of product variety that encompasses both vertical and horizontal differentiation notions, and work with a "democratic" dataset that pertains to consumers' actual wine purchases for U.S. home consumption. In particular, our study uses nearly one million observations about

U.S.-produced bottles of wine, encompassing 38 varietals, 3,939 brands, and 80 distinct geographic origins.

The second distinguishing feature of this study is that we use the estimated hedonic equation to provide empirical evidence on the extent to which wine characteristics can explain wine market prices. For that purpose, we introduce to the wine hedonic analysis a metric based on the Shapely value, from co-operative game theory, that permits us to shed light on the relative importance of broad sets of the explanatory variables included in the model as wine price predictors.

We find that, even after accounting for firms' brands, wine types, varietals, and other control variables likely to affect retail prices, label information about the geographic origin of U.S. wines carries considerable explanatory power. In particular, AVA labels are associated with non-negligible price premia, relative to an undefined U.S. origin. In quantitative terms, using the Shapley value metric, we find that over 70% of U.S. retail wine prices are accounted for by individual wines' own brands. This finding is consistent with the basic economics of markets for experience goods: branding can be effective in providing credible quality signals and fostering repeat purchases. Next to that, however, it is information about the geographic origin that matters most, with the Shapley values suggesting the contribution of this set of variables ranges from 12 to 14%. In particular, it seems that the incidence of geographic origin information is twice as large as that provided by varietals.

This paper is organized as follows. Section II briefly reviews the relevant literature on wine price determinants in the hedonic price framework. Section III discusses the data sources and descriptive statistics. Section IV presents the hedonic price model. Section V reports the estimation results, including estimated implicit prices of various attributes and the Shapley value characterization of the relative importance of a set of attributes. Section VI concludes.

II. Related literature

Wine attributes of interest to consumers include so-called intrinsic characteristics, such as grape variety, region, and vintages. Extrinsic characteristics, such as expert ratings and tasting notes, may also influence consumers' appreciation of the product. Many empirical studies apply the hedonic method to investigate the importance of these attributes in explaining prices. Our discussion of related literature is necessarily brief. For more detailed and comprehensive surveys, see Oczkowski and Doucouliagos (2015) and Outreville and Le Fur (2020).

Using the Bordeaux region appellation of origin as an indicator for collective reputation, Landon and Smith (1997, 1998) present the first empirical analyses measuring the impact of reputation on wine prices. The authors estimated hedonic price functions for Bordeaux wine and found that geographic origin indication has a larger impact on consumer willingness to pay than grape variety. Subsequent studies underscore the role of geographic origin in wines. Schamel (2002) for California wines, Schamel and Anderson (2003) for Australian and New Zealand wines, Roma, Di Martino, and Perrone (2013) and Levaggi and Brentari (2014) for Italian wines, and Troncoso and Aguirre (2006) for Chilean wines, find a significant impact of geographic origin on price. Other studies also look at the importance of geographic origin relative to the grape variety. Steiner (2004) for Australia, and Costanigro, McCluskey, and Mittelhammer (2007) for the North American region, find geographic origin as a more significant determinant of wine price than grape varietals. Kwon, Lee, and Sumner (2008) show that grape variety and appellation interaction significantly influence wine prices for California wines even after controlling for vintage and tasting scores.

Wine's sensory characteristics seem to be less important to consumers compared to label information such as varietal or geographic origin (Oczkowski, 1994; Combris, Lecocq, and Visser, 1997; Cardebat and Figuet, 2004). A possible explanation is that most consumers do not understand or have limited knowledge about wine's technical aspects, such as sugar, tannins, and other sensory attributes. Thus, sensory attributes may not influence price from the demand side (Outreville and Le Fur, 2020; Combris, Lecocq, and Visser, 1997).

Hedonic price studies on wine also look at the relative importance of individual and collective reputation in explaining wine prices. Costanigro, McCluskey, and Goemans (2010) present the idea of nested reputation and jointly analyze the effects of product, firm, and collective reputation on the market prices of California wine. These authors conclude that the relative importance of reputation changes with product price—the reputation premia shifts from collective to specific names as product price increases. Schamel (2009), using data on 27 wine-growing regions, tests whether individual brands with a strong quality reputation rely less on the region's reputation and finds mixed evidence for this hypothesis.

Experimental methods have also been applied to assess consumers' valuation of wine attributes and prices for purchase decisions. Through a choice experiment, Lockshin et al. (2006) find that information about geographic origin increases retail sales of wine; however, the effect differs between high and low engagement customers. Gustafson (2011), using a lab experiment, finds that appellation is a highly-valued attribute for California wine consumers. In another lab experiment, Gustafson, Lybbert, and Sumner (2016) analyze the effect of consumers' knowledge on the ability to interpret wine information. These authors find that high-knowledge consumers do not value wine differently from low-knowledge consumers but updated their bid considerably when told about attribute information.

III. Data sources and descriptive statistics

This study uses Nielsen's consumer panel data collected for the U.S. market from 2007 to 2019. Household-based homescan data are obtained from a panel of approximately 60,000 households. The nationally representative sample of panelists is sampled from all states and major markets and is demographically balanced. Nielsen uses a stratified, proportionate sample that represents the entire United States into 61 geographic areas, including 52 major markets and nine census divisions. We restrict the analysis in this chapter to the major Scantrack markets.[2] Nielsen-defined major

[2]Nielsen's coding system for Scantrack markets changed in 2016. Until then, Nielsen distinguished three markets for New York City (urban, exurban, and suburban New York City). From 2016 onwards, Nielsen includes just the New York City market. For consistency, we combined the three NY regions for all years before 2016, and thus the analysis covers 50 distinct Scantrack markets.

Table 1. Sample composition and price distribution, by wine type

			Price ($ per 750 ml bottle)			
	Percent share (quantity)	Number of brands	Mean	1st pct	50th pct	99th pct
Red	54.79	3294	9.22	2.26	8.46	29.28
Rosé	7.80	601	5.87	2.20	5.38	14.99
White	37.41	2391	8.05	2.26	7.39	21.77

Note: Summary statistics are weighted using Nielsen projection factor.
Source: Nielsen Consumer Panel Data.

Scantrack markets are similar to the metropolitan statistical area used in the U.S. census.

Nielsen panelists continually provide information about each shopping trip. The information provided for each trip includes the date, retailer code, store code, and total dollars spent on that trip. Panelists also provide detailed transaction information for each item purchased within a trip, including the UPC, quantity, and deals or coupons if used. If a transaction involves coupons, then households also record the amount of the coupon. Further, for each UPC, the dataset also provides a shorthand description. A major component of our data work was to extract relevant information concerning wine type, varietal, and geographic origin of the purchased wine from the text of UPC descriptions. Appendix B provides more details on this process.

Nielsen classifies wine products into 12 different modules. Table A1 in the Appendix reports some descriptive data. In this study, we focus on the largest of these modules, which pertains to domestic dry table wine (which accounts for about 65% of the volume of total at-home U.S. wine consumption). Furthermore, given the focus of our analysis, we concentrate on wine sold in standard 750 ml bottles (thus excluding wine sold in bulk containers). All prices are expressed in $/bottle and are deflated by the Consumer Price Index (2019=1).

We distinguish three types of wine: red, white, and rosé. Table 1 reports some descriptive statistics of our sample, which includes nearly one million bottles, 55% of which are red wine, 37% are white wine, and about 8% are rosé. The price distribution indicates that red wine commands a premium relative to both white and rosé, on average. The overall price distribution is somewhat right-skewed.

Table 2 reports descriptive statistics for red, white, and rosé varietals.[3] We capture information on 17 red, 13 white, and 8 rosé wine varietals from the data. Additionally, we aggregate any varietal with fewer than 50 observations in the data under the "Other" category. In terms of quantity, Cabernet Sauvignon, Chardonnay, and Zinfandel rosé are the top-selling varietals among red, white, and rosé wines,

[3]The "varietals" defined in this paper do not always represent unique types of grapes. We find many blended varietals in the data, such as Cabernet-Syrah, Merlot-Malbec, Chardonnay-Chenin Blanc, and Chardonnay-Pinotage. Most of these blend varietals have a very small share of the total quantity. Thus, we aggregate blend varietals. For instance, Cabernet Blend includes blends of Cabernet with Merlot, Malbec, and/or Syrah.

Table 2. Sample composition and prices, by varietals

	% Share	No. of brands	Mean price		% Share	No. of brands	Mean price
Red Wine				White Wine			
Barbera	0.02	23	12.58	Chardonnay	18.07	1,457	8.59
Cabernet Franc	0.02	51	16.29	Chardonnay Blend	0.03	14	5.10
Cabernet Sauvignon	17.87	1,425	9.37	Chenin Blanc	0.47	45	6.72
Cabernet Blend	0.16	53	10.40	Gewurztraminer	0.77	81	7.98
Concord	0.06	22	7.90	Moscato White	2.54	175	5.99
Grenache Blend	0.01	22	12.75	Muscadine White	0.04	17	8.08
Malbec	0.09	39	9.14	Other White	2.82	687	9.50
Merlot	11.60	805	7.52	Pinot Grigio	5.37	487	6.55
Moscato Red	0.37	21	6.34	Pinot Blanc	0.02	20	14.09
Muscadine Red	0.01	5	8.24	Riesling	3.32	316	8.58
Norton	0.01	9	9.48	Sauvignon Blanc	3.84	490	7.95
Other Red	12.24	1,581	9.62	Semillon	0.01	7	10.91
Petite Syrah	0.55	133	10.81	Viognier	0.11	88	11.54
Pinot Noir	5.72	888	11.37				
Syrah	2.69	324	5.51				
Tempranillo	0.01	12	14.35				
Zinfandel	3.36	502	11.43				

(*Continued*)

Table 2. (*Continued.*)

	% Share	No. of brands	Mean price		% Share	No. of brands	Mean price
Rosé Wine							
Cabernet Rosé	0.01	4	6.71	Pinot Grigio Rosé	0.02	6	7.88
Grenache Rosé	0.01	12	9.11	Pinot Noir Rosé	0.09	49	12.61
Merlot Rosé	0.61	17	5.75	Pink Moscato	0.61	22	5.65
Other Rosé	1.10	418	8.68	Zinfandel Rosé	5.36	136	5.13

Notes: The % share is computed in volume (quantity) terms, weighted using Nielsen projection factors. Mean price is $/bottle. Cabernet Blend includes Cabernet-Malbec, Cabernet-Merlot, Cabernet-Merlot-Cabernet Franc, Cabernet Syrah. Chardonnay blend includes Chardonnay-Chenin Blanc, Chardonnay-Pinot Noir, Chardonnay-Pinotage, Chardonnay-Semillon, Chardonnay-Viognier. Other White and Other Red include all other entries for which we do not have specific varietal information in the dataset. Other Red also includes red blends.
Source: Nielsen Consumer Panel Data.

respectively. In terms of average price, Cabernet Franc, Tempranillo, Pinot Blanc, Viognier, and Pinot Noire rosé fetch a premium relative to other varietals.

We found 3,939 unique brands in the sample over the 13-year period. The average number of brands in a year is about 1,500. The top 50 brands account for about 68% of the total quantity observed in the sample. Table A2 in the Appendix lists the top 50 brands in total quantity in alphabetical order. Based on the information found in the dataset, we can broadly classify geographic information on wine products into three categories: (a) wines with AVA or county names; (b) wines with state appellation of origin; and (c) wines with no specific geographic origin information, referred to in this paper as "U.S. generic." AVAs are administered by the federal government and provide the most selective tool of geographic differentiation used by U.S. wine producers. Currently, there are 261 approved AVAs in the United States, 143 of which are in California. Under U.S. law, a viticulture area for American wine is "a delimited grape-growing region having distinguishing features, a name and delineated boundary." AVAs allow vintners and consumers to attribute a given quality, reputation, or other wine characteristics made from grapes grown in an area to its geographic origin.

The Alcohol and Tobacco Trade and Tax Bureau (TTB) is responsible for designating and reviewing all petitions to establish a new or expand an existing AVA. Once the TTB approves an AVA, a producer can include an AVA label if at least 85% of the volume of wine is derived from grapes grown in the named viticulture area.[4] Augusta, Missouri, was the first wine-growing region approved as an AVA by the TTB in June 1980. Alternatively, producers can use the name of the state or county on their wine bottles if 75% or more of the volume of wine is derived from grapes grown in the named state or county (California and Washington have stricter requirements). In our sample, we find that 23.5% of domestic wine bottles contain AVAs or county appellations, 50% have a state appellation (most of it pertains to the California state appellation), and 26.4% do not claim any particular geographic origin.

Our data capture information on 115 AVAs, but any AVA with fewer than 50 observations is aggregated under the broader region (AVA or county) that it is part of, or is grouped under the "Other AVAs" category. In the end, we are left with information on 81 distinct geographic origins, including 65 AVAs or counties, 15 state appellations, and one U.S. generic group. Table 3 presents the percent share in total quantity, brand presence, and average price (dollars per bottle) for each selected geographic origin.[5]

Along with these characteristics, earlier work has shown that vintage and alcohol content are also important attributes of wine prices. However, this information is not available in homescan data. Recent studies that use datasets similar to ours, for other regions, also do not have vintage information (Carew, Florkowski, and Meng, 2017).

[4]The wine must also be finished in the state(s) where the viticulture area is located. The TTB includes an exception for cellar treatment and blending, neither of which has to occur in the labeled area.

[5]The geographic origin that we extract from the UPC label comes in a variety of short names, and some discretion has to be used to interpret and group them. Some of the entries in Table 3 are single AVAs (e.g., Mendocino or Napa Valley), some are individual counties (e.g., Amador County), and others may combine a county and small AVAs in that county that are not considered individually (e.g., Lake County or Sonoma County). The aggregate "Other AVAs" only includes AVAs not otherwise considered, and county labels not considered individually are grouped in their own state appellation.

https://doi.org/10.1017/jwe.2022.23 Published online by Cambridge University Press

Table 3. Sample composition and prices, by geographic origin

Geographic origin	% Share	No. of brands	Mean price	Geographic origin	% Share	No. of brands	Mean price
AVA / County				Rogue Valley	0.01	11	13.52
Alexander Valley	0.34	93	16.31	Russian River Valley	0.33	157	16.69
Ancient Lakes	0.01	3	9.88	Rutherford	0.08	22	20.34
Altus	0.00	3	7.51	Santa Lucia Highlands	0.03	31	18.06
Amador County	0.08	24	13.24	San Antonio Valley	0.04	3	10.62
Anderson Valley	0.01	15	25.69	San Bernabe	0.01	3	10.18
Arroyo Seco	0.08	24	14.35	Santa Barbara County	0.42	94	9.61
Augusta	0.03	2	7.38	Santa Maria Valley	0.09	24	29.11
Carneros	0.21	73	20.61	Santa Ynes Valley	0.01	12	15.18
Central Coast	3.87	211	9.86	Shenandoah Valley	0.01	3	10.76
Chalk Hill	0.02	7	18.14	Sierra Foothills	0.02	30	15.37
Clarksburg	0.05	18	12.64	Snake River Valley	0.04	6	7.16
Columbia Valley	4.42	249	9.51	Sonoma County	2.03	214	12.56
Dry Creek Valley	0.16	66	15.54	Sonoma Coast	0.48	80	17.50
Dunnigan Hills	0.02	6	11.97	Sonoma Valley	0.06	28	19.49
Eagle Peak	0.05	1	6.73	South Coast	0.01	2	10.92
Edna Valley	0.52	32	10.90	St. Helena	0.00	11	16.34
El Dorado	0.01	8	14.61	Sta. Rita Hills	0.01	18	17.87
Eola Hills	0.04	5	11.34	Temecula Valley	0.01	10	15.82
Finger Lakes	0.08	16	11.40	Wahluke Slope	0.05	14	11.96
Guenoc	0.11	2	9.06	Walla Walla Valley	0.06	19	16.98

High Valley	0.02	8	15.81	Willamette Valley	0.26	136	16.73
Horse Heaven Hills	0.28	22	12.38	Yadkin Valley	0.01	5	12.74
Knights Valley	0.05	12	23.53	Yakima Valley	0.06	42	12.56
Lake County	0.18	45	11.22	State			
Lake Erie	0.02	11	9.33	Arkansas	0.02	2	7.11
Lake Michigan Shore	0.02	7	8.46	California	47.91	1,170	6.94
Leelanau Peninsula	0.01	9	14.22	Idaho	0.01	4	9.64
Livermore Valley	0.17	20	12.86	Illinois	0.02	4	7.30
Lodi	1.55	195	12.08	Indiana	0.01	5	9.82
Mendocino	0.39	93	12.64	Missouri	0.06	14	7.74
Monterey County	2.06	177	11.74	Nebraska	0.01	6	16.37
Napa County	0.33	63	11.71	New York	0.09	22	9.09
Napa Valley	1.97	421	15.43	North Carolina	0.09	19	9.10
North Coast	1.51	131	12.03	Ohio	0.02	8	10.25
Oakville	0.02	17	35.35	Oregon	0.25	85	14.15
Ohio River Valley	0.04	5	10.49	Other States	0.04	61	13.68
Other AVAs	0.03	59	15.56	Pennsylvania	0.01	10	12.04
Paso Robles	0.67	173	13.17	Texas	0.06	29	9.78
Rattlesnake Hills	0.01	9	10.93	Virginia	0.03	20	14.98
Red Hills Lake County	0.02	12	13.32	Washington	1.47	98	9.41
Redwood Valley	0.01	8	15.50	United States	26.37	1,495	8.07

Notes: The % share is computed in volume (quantity) terms and weighted using Nielsen projection factors. Mean price is $/bottle.
Source: Nielsen Consumer Panel Data.

We use data at the household-trip-UPC level to model the hedonic relationship between wine price and attributes. There are 997,521 observations over the 13-year period from 2007 to 2019. The data includes wine purchase information for over 13,763 unique UPCs over the 13-year period, with an average of about 4,300 distinct UPCs each year. We also include some variables capturing the market characteristics along with wine attributes, which can explain some price variation across markets. Next we will discuss each one.

(1) Channel type

Nielsen consumer panel data also provides information on the channel type from which households purchase wine. Nielsen classifies different channels into 66 mutually exclusive categories. The top nine channel types capture about 96% of the total quantity purchased in the sample. Therefore, we include the top nine channel types and aggregate all other channels into one category. Information on channel type is included in the model to capture variations in price across the retail channel. Table 4 presents the percent share of the top nine channels, the presence of brands, and the average price for a bottle of wine across these channels.

(2) Retail density

We also capture retail density at the county level to account for price variation across the market. We collect data from County Business Pattern (CBP) at the county level to measure retail density. CBP releases data annually and provides economic data by industry, including the count of establishments and employment level, among other information. CBP follows and provides data from the North American Industry Classification System (NAICS). Under NAICS, retail trade for food and beverage

Table 4. Sample composition and price distribution, by retailing channel type

	% Share	No. of brands	Mean price
Beverage store	0.55	865	11.78
Discount store	8.84	743	5.96
Drug store	2.89	444	6.48
Grocery	57.21	3,004	7.85
Home furnishings	0.62	580	10.82
Liquor store	16.43	2,730	11.23
Military store	0.52	389	8.95
Online shopping	0.54	621	10.35
Warehouse club	9.29	1,079	10.51
All other channels	3.10	1,909	9.58

Notes: The % share is computed in volume (quantity) terms and weighted using Nielsen projection factors. Mean price is in $/bottle.
Source: Nielsen Consumer Panel Data.

stores is further classified under three sub-categories: (a) grocery stores, sub-classified as supermarkets and other grocery (except convenience stores) or convenience stores; (b) specialty food stores, sub-classified as meat markets, fish and seafood markets, fruit and vegetable markets, or other specialty food stores; and (c) beer, wine, and liquor stores. To count the number of establishments selling wine, we include establishments under categories (a) and (c). We use county-level population and land area to calculate the retail density normalized by spatial area and population. We use American Community Survey (ACS) annual data by the U.S. census to collect county-level population and U.S. census 2010 data to collect the "land area" of a given county. The county-level population from ACS is collected from the IPUMS NHGIS database.[6] We define "population-adjusted retail density" as the number of establishments per 1,000 residents, and "spatial area adjusted retail density" as the number of establishments per square mile.

(3) State excise tax on wine

The excise tax on wine levied by the state varies a lot across the United States. Along with a three-tier distribution system,[7] states also impose constraints on the distribution and sales of alcoholic beverages and maintain distribution franchise laws (Santiago and Sykuta, 2016). Young and Bielińska-Kwapisz (2002) show that excise taxes on alcoholic drinks lead to increased alcohol retail prices. Thus, we also control for a state excise tax to capture some variation in wine prices across different states.

We collect wine excise tax data from the Tax Foundation and Federation of Tax Administrators (FTA). States apply varying excise tax rates based on wine type and alcohol content. In 2020, excise taxes differed significantly across states, with the highest in Alaska at $2.50 per gallon, and Pennsylvania, Utah, and Wyoming had no excise tax. These three states with zero excise tax are here referred to as "state control" states because the state government essentially controls all retail wine sales. There is no explicit excise tax on wine in these states, and revenue is generated through various other taxes, fees, price mark-ups, and net profits (FTA). Under the federal code and in some states, sparkling wine has a different excise tax than other wines.[8] In this study, we consider two categories for excise tax: (a) base excise tax for all wine other than sparkling wine; and (b) excise tax on sparkling wine. In states for which we do not have information about the excise tax on sparkling wine, we consider the base excise tax for sparkling wine. We also account for "state control" states in the model and capture the retail density and excise tax information for only those markets that do not fall under "state control" systems.

IV. The hedonic price model

In a differentiated product market, there are clear modeling advantages to framing consumers' demand in the characteristics space, following the pioneering work of

[6]IPUMS USA, University of Minnesota, www.ipums.org.

[7]The United States follows a three-tier distribution system for alcohol, where producers sell to wholesalers or distributors, distributors sell to retailers, and retailers sell to consumers.

[8]In the FTA data, we find additional excise tax information for sparkling wine for around 13 states.

Gorman (1956) and Lancaster (1966). Traded products are thought of as bundles of characteristics, which is what consumers ultimately care about. There are no markets for the characteristics per se, but the prices of products that are actually bought and sold implicitly define the prices of the characteristics included in them. Rosen's (1974) seminal paper formalized this insight in the context of a market with a continuum of products and perfect competition. In this setting, the price of a product turns out to be a function of the product's content of characteristics. Consumers are heterogeneous with respect to income and/or preferences for characteristics, and profit-maximizing firms face a convex cost of supplying those characteristics. In Rosen's (1974) competitive equilibrium, there exists a price function $p(z)$ such that a good that embeds a vector of characteristics z^i commands a price $p(z^i)$. This function, termed the hedonic price function, is the envelope of both consumers' indifference curves (bid functions) and the firms' iso-profit curves (offer functions). Thus, the hedonic price function represents an equilibrium relation that captures the market's valuation of products' characteristics that have economic relevance.

It is important to underscore that the hedonic price function describes an equilibrium outcome. Rosen's (1974) original characterization presumed pure competition, but clearly, an equilibrium price relation linking a product's price to its characteristics also applies to non-competitive markets (Pakes, 2003; Bajari and Benkard, 2005; Nesheim, 2008). As such, the function $p(z)$ reflects both demand factors pertaining to consumers' preferences (e.g., consumers' willingness to pay for individual attributes) as well as supply-side factors (including production costs associated with characteristics and/or market power markups). Disentangling the separate impacts of demand and supply factors is, in general, a difficult matter. The canonical hedonic price function framework envisioned by Rosen (1974) presumes two empirical stages. The first stage uncovers the structure of the hedonic price function by regressing the prices of traded products against their characteristics. The empirical relation $\hat{p}(z)$ thus obtained defines the individual characteristics' implicit prices $\partial\hat{p}(z)/\partial z_k$. Such marginal prices naturally reflect the underlying demand and supply conditions, the effects of which could conceivably be uncovered in a second stage, where the estimated characteristics' implicit prices are the dependent variables.

A large literature articulates the drawbacks of such a two-stage approach. In particular, the identification issues that arise with the second stage are daunting (Brown and Rosen, 1982; Ekeland, Heckman, and Nesheim, 2002), and standard supply-side shifter instruments do not work (Bartik, 1987; Epple, 1987). As a result, empirical analyses of hedonic price functions are often confined to the first stage. Researchers wishing to disentangle demand and supply determinants on the equilibrium prices of differentiated products typically pursue more structural approaches, which are, by construction, amenable to welfare assessments and the study of counterfactual policies (Gandhi and Nevo, 2021).

This paper follows many other studies by restricting attention to the first-stage estimation of the hedonic price function. Because, as articulated in the foregoing, the hedonic price function characterizes equilibrium outcomes, and given that we look at actual retail prices over a large set of market conditions, the determinants of wine prices that we consider include a few variables related to the retail market, in addition to the wines' own characteristics.

A. Hedonic price function

We express the price of wine as a function of the characteristics of interest—wine type, varietals, brand name, and geographic origin—and of market features. Among the latter, we include excise tax, retail density, state distribution rule, the distribution channel of the purchased wine, as well as fixed effects that capture systematic spatial and temporal price effects common across wines. Consistent with much of the previous work on hedonics in wine research reviewed earlier (e.g., Oczkowski, 1994; Combris, Lecocq, and Visser, 1997; Carew, Florkowski, and Meng, 2017), we adopt the semi-log parameterization.[9] That is, the hedonic price equation that is estimated is expressed as:

$$\ln p_{ji} = \alpha + \mathbf{z}_j'\boldsymbol{\beta} + \mathbf{x}_{ji}'\boldsymbol{\lambda} + \gamma(1 - D_s)T_{st} + \phi D_s + \eta(1 - D_s)R_{ct} + \xi_b + \xi_m + \xi_q + \xi_t + \varepsilon_{ji}, \tag{1}$$

where p_{ji} is the price of wine product j (at the UPC level) observed at purchase occasion i; and, each purchase occasion pertains to a specific Nielsen-defined major market $m = m[i]$ and occurred in a specific year $t = t[i]$. On the right-hand-side, $\mathbf{z}_j$ is a vector of wine attributes, which includes wine type, varietal, and geographic origin; $\mathbf{x}_{ji}$ is a vector of dummies that control for the retail store type in which product j was bought at purchase occasion i; D_s is a dummy variable that denotes markets that falls in a state that has total control over wine distribution (here, $s = s[i]$ indicates the state where purchase i took place); T_{st} denotes the state excise tax in the year t; and, R_{ct} is the measure of retail density for county c (where the purchase is made) in year t. Further controls are provided by a rich set of fixed effects: ξ_b is the brand fixed effect (where $b = b[j]$ denotes the brand associated with wine product j); ξ_m is the market fixed effect; and, ξ_q and ξ_t are, respectively, the quarter and the year fixed effects. Finally, ε_{ji} is the error term, assumed to be identically and independently distributed.

V. Results

Table 5 summarizes the variables included in the hedonic price regression and provides short descriptions. The estimation results are reported in Table 6. Whereas the entire set of estimates will be used for the Shapley value metrics in the next section, here we focus on entries for varietals and geographic origins that have a solid representation in the estimation sample: specifically, at least 0.05% in quantity share (roughly speaking, about 500 bottles over the 13-year period) and at least 10 different brands. The first three columns of results in Table 6 pertain to the hedonic price model as presented in Equation (1), whereas the last three columns are from an expanded model that includes interaction effects between varietals and geographic origin variables.

In the hedonic price equation, we considered four sets of explanatory variables that relate to wine's intrinsic attributes—wine type, varietals, geographic origin, and

[9]The choice of this functional form appears quite consistent with our data. See Table A3 in the Appendix for some evidence from a Box-Cox specification.

Table 5. Variables in the hedonic regression

Log of price	The natural logarithm of price is the dependent variable. Prices are in $ per 750 ml bottle, deflated by the Consumer Price Index (2019 = 1)
Wine type	2 indicator variables coding White and Rosé (Red wine is the reference)
Varietals	35 indicator variables: 16 for red wine (Merlot is the base), 12 for white wine (Pinot Grigio is the base), and 7 for rosé wine (Zinfandel rosé is the reference)
Geographic origin	80 indicator variables: 65 AVAs and 15 state appellations
Retail distribution channels	9 indicator variables (grocery stores are the base reference)
State control	1 indicator variable, for states with zero excise tax (includes Pennsylvania, Utah, and Wyoming)
Excise tax	1 continuous variable, for states with excise taxes on wine
Retail density	2 continuous variables, measuring the density of retail stores in terms of population and in terms of geographic area
Quarter fixed effects	3 indicator variables
Year fixed effects	12 indicator variables
Market fixed effects	49 indicator variables
Brand fixed effects	3,938 indicator variables

Note: This table provides a summary of the variables included in the hedonic regression.

brand. In addition, because we are fitting equilibrium retail prices, we also include a set of variables pertaining to retailing conditions, as well as year, quarter, and market fixed effects. The interpretation of these last two sets of variables, of course, differs from the intrinsic wine attributes, but it is important to include them as controls for correct inference about wine's intrinsic attributes.

A. Baseline hedonic model

Because the semilog functional form is used, the coefficients reported in Table 6 are approximately related to the percent premium (or discount) of each attribute relative to the base reference. Specifically, if β_k is the estimated coefficient associated with an attribute coded by a dummy indicator variable, the corresponding implicit price, expressed as a percent of the reference wine price, is computed as $100[\exp(\beta_k) - 1]$ (Halvorsen and Palmquist, 1980). This percent implicit price is also reported in Table 6. Because of the large number of observations (nearly one million), most estimated coefficients are significantly different from zero at conventional significance levels.

Concerning wine type, we see that red wine (the base reference) carries a premium of about 4%, relative to both white and rosé wine. For varietals, we have a different reference base for each type. Relative to the reference for red wine (Merlot), we find that Cabernet Sauvignon, Malbec, Petite Syrah, Pinot Noir, and Zinfandel all carry a moderate premium, the largest one commanded by Pinot Noir (7.7%). Among white varietals, relative to the chosen reference base (Pinot Grigio), the wines with the largest discounts are Chenin Blanc (–11.3%) and Riesling (–6.3%),

Table 6. Hedonic price regressions results

	Baseline hedonic model			Model with interaction effects		
	Coefficient	p-value	Implicit price (%)	Calculated coefficient	p-value	Implicit price (%)
Wine type (base: Red)						
Rose	−0.048	(0.000)	−4.7	−0.037	(0.000)	−3.6
White	−0.039	(0.000)	−3.8	−0.074	(0.000)	−7.1
Red Varietals (base: Merlot)						
Cabernet Sauvignon	0.045	(0.000)	4.6	0.049	(0.000)	5.0
Cabernet Blend	−0.037	(0.000)	−3.6	−0.083	(0.000)	−8.0
Malbec	0.063	(0.000)	6.5	0.047	(0.000)	4.8
Moscato Red	−0.039	(0.000)	−3.8	−0.021	(0.000)	−2.1
Petite Syrah	0.045	(0.000)	4.6	0.043	(0.000)	4.4
Pinot Noir	0.074	(0.000)	7.7	0.084	(0.000)	8.8
Syrah	−0.042	(0.000)	−4.1	−0.036	(0.000)	−3.5
Zinfandel	0.029	(0.000)	2.9	0.052	(0.000)	5.3
Other Red	0.005	(0.000)	0.5	0.016	(0.000)	1.6
White Varietals (base: Pinot Grigio)						
Chardonnay	−0.0043	(0.000)	−0.4	0.025	(0.000)	2.5
Chenin Blanc	−0.12	(0.000)	−11.3	−0.083	(0.000)	−8.0
Gewurztraminer	−0.011	(0.001)	−1.1	0.007	(0.174)	0.7
Moscato White	−0.0042	(0.017)	−0.4	0.042	(0.000)	4.3

(*Continued*)

https://doi.org/10.1017/jwe.2022.23 Published online by Cambridge University Press

Table 6. (*Continued.*)

	Baseline hedonic model			Model with interaction effects		
	Coefficient	p-value	Implicit price (%)	Calculated coefficient	p-value	Implicit price (%)
Riesling	−0.065	(0.000)	−6.3	−0.089	(0.000)	−8.5
Sauvignon Blanc	−0.020	(0.000)	−2.0	0.015	(0.000)	1.5
Viognier	0.049	(0.000)	5.0	0.092	(0.000)	9.6
Other White	−0.013	(0.000)	−1.3	0.010	(0.004)	1.0
Rosé Varietals (base: Zinfandel Rosé)						
Pinot Noire Rosé	−0.15	(0.000)	−13.9	−0.147	(0.000)	−13.7
Moscato Rosé	−0.029	(0.000)	−2.9	−0.068	(0.000)	−6.6
Merlot Rosé	−0.13	(0.000)	−12.2	−0.197	(0.000)	−17.9
Other Rosé	−0.00079	(0.793)	−0.1	−0.016	(0.000)	−1.6
Region (base: generic United States)						
AVA / County						
Alexander Valley	0.35	(0.000)	41.9	0.37	(0.000)	44.8
Amador County	0.21	(0.000)	23.4	0.25	(0.000)	28.4
Arroyo Seco	0.37	(0.000)	44.8	0.36	(0.000)	43.3
Carneros	0.27	(0.000)	31.0	0.29	(0.000)	33.6
Central Coast	0.11	(0.000)	11.6	0.11	(0.000)	11.6
Clarksburg	0.16	(0.000)	17.4	0.17	(0.000)	18.5
Columbia Valley	0.00041	(0.897)	0.0	−0.02	(0.000)	−2.0

(*Continued*)

https://doi.org/10.1017/jwe.2022.23 Published online by Cambridge University Press

Table 6. (*Continued.*)

	Baseline hedonic model			Model with interaction effects		
	Coefficient	p-value	Implicit price (%)	Calculated coefficient	p-value	Implicit price (%)
Dry Creek Valley	0.39	(0.000)	47.7	0.37	(0.000)	44.8
Eagle Peak	−0.11	(0.000)	−10.4	−0.09	(0.000)	−8.6
Edna Valley	−0.055	(0.000)	−5.4	−0.05	(0.000)	−4.9
Finger Lakes	−0.061	(0.000)	−5.9	0.005	(0.675)	0.5
Horse Heaven Hills	0.22	(0.000)	24.6	0.31	(0.000)	36.3
Knights Valley	1.16	(0.000)	219.0	1.15	(0.000)	215.8
Lake County	0.020	(0.036)	2.0	0.062	(0.000)	6.4
Livermore Valley	0.23	(0.000)	25.9	0.18	(0.000)	19.7
Lodi	0.055	(0.000)	5.7	0.05	(0.000)	5.1
Mendocino	0.029	(0.000)	2.9	0.07	(0.000)	7.3
Monterey County	0.092	(0.000)	9.6	0.11	(0.000)	11.6
Napa County	0.17	(0.000)	18.5	0.23	(0.000)	25.9
Napa Valley	0.25	(0.000)	28.4	0.37	(0.000)	44.8
North Coast	0.069	(0.000)	7.1	0.08	(0.000)	8.3
Paso Robles	0.17	(0.000)	18.5	0.17	(0.000)	18.5
Russian River Valley	0.37	(0.000)	44.8	0.39	(0.000)	47.7
Rutherford	1.13	(0.000)	209.6	1.10	(0.000)	200.4
Santa Barbara County	0.0073	(0.261)	0.7	0.012	(0.198)	1.2

(*Continued*)

Table 6. (*Continued.*)

	Baseline hedonic model			Model with interaction effects		
	Coefficient	p-value	Implicit price (%)	Calculated coefficient	p-value	Implicit price (%)
Santa Maria Valley	0.36	(0.000)	43.3	0.35	(0.000)	41.9
Sonoma County	0.24	(0.000)	27.1	0.22	(0.000)	24.6
Sonoma Coast	0.20	(0.000)	22.1	0.17	(0.000)	18.5
Sonoma Valley	0.54	(0.000)	71.6	0.51	(0.000)	66.5
Walla Walla Valley	0.11	(0.000)	11.6	0.108	(0.000)	11.4
Willamette Valley	0.32	(0.000)	37.7	0.309	(0.000)	36.2
Yakima Valley	−0.0052	(0.842)	−0.5	0.080	(0.052)	8.3
State						
California	−0.014	(0.000)	−1.4	−0.021	(0.000)	−2.1
Missouri	0.020	(0.030)	2.0	−0.004	(0.616)	−0.4
New York	0.048	(0.014)	4.9	0.050	(0.004)	5.1
North Carolina	0.044	(0.000)	4.5	0.021	(0.009)	2.1
Oregon	0.13	(0.000)	13.9	0.056	(0.016)	5.8
Texas	0.058	(0.000)	6.0	0.068	(0.000)	7.0
Washington	−0.00097	(0.867)	−0.1	0.008	(0.161)	0.8

(*Continued*)

Table 6. (*Continued.*)

	Baseline hedonic model			Model with interaction effects		
	Coefficient	p-value	Implicit price (%)	Calculated coefficient	p-value	Implicit price (%)
Channel (base: Grocery store)						
All other channels	0.058	(0.000)	6.0	0.058	(0.000)	6.0
Beverage store	0.050	(0.000)	5.1	0.049	(0.000)	5.0
Discount store	−0.0037	(0.000)	−0.4	−0.0038	(0.000)	−0.4
Drug store	−0.024	(0.000)	−2.4	−0.022	(0.000)	−2.2
Home furnishings	0.0099	(0.000)	1.0	0.011	(0.000)	1.1
Liquor store	0.053	(0.000)	5.4	0.051	(0.000)	5.2
Military store	−0.0096	(0.000)	−1.0	−0.0098	(0.000)	−1.0
Online shopping	0.035	(0.000)	3.6	0.033	(0.000)	3.4
Warehouse club	−0.050	(0.000)	−4.9	−0.057	(0.000)	−5.5
Observations	996,956			996,956		
R^2	0.80			0.81		

Notes: There are 997,521 observations in total. However, 565 observations that are singletons are dropped when estimating fixed effect regression. The fixed effect regression is estimated using the "REGHDFE" command in Stata. The "calculated coefficients" reported for the "interaction" model pertain to the total effect evaluated at the appropriate conditional mean (see text for more details). The standard error for the interaction model is calculated using the delta method.

whereas Viognier carries a premium of 5%. Chardonnay is essentially equivalent to the Pinot Grigio reference. All rosé varietals included appear to have a discount relative to the base varietal (Zinfandel rosé).

Looking at the geographic origin, it is apparent that most AVA wines carry a considerable price premium relative to the base reference (unspecified U.S. origin). Some selective AVAs like Knight Valley and Rutherford show the largest price premia by far (exceeding 200%). Other well-known AVAs with substantial price premia include Alexander Valley (41.9%), Carneros (31%), Dry Creek Valley (47.7%), Russian River Valley (44.4%), Santa Maria Valley (43.3%), Sonoma Valley (71.6%), and Willamette Valley (37.7%). The price premium associated with Napa labels, by contrast, is somewhat smaller (28.4% for Napa Valley and 18.5% for Napa County). Among the large and well-known AVAs, Columbia Valley is the only one to show a zero price difference relative to the reference (generic U.S. origin).

In interpreting these results, it is important to recall that these are equilibrium price differentials. In addition to consumers' willingness to pay, they also reflect supply conditions. Furthermore, the estimated price differentials for attributes, such as geographic origin, are conditional on all other variables in the model. Chief among the latter are brands. Indeed, if a firm succeeded in capturing the full set of price-relevant information by means of a brand label, the scope for characteristics to explain price would be void.

As for state appellation, among the large wine-producing states, only Oregon carries a sizeable price premium (13.9%), whereas California and Washington show negligible (and negative) price differentials relative to the reference base. Smaller wine-producing states such as Missouri, New York, North Carolina, and Texas all show moderate and positive price premia relative to the reference (generic U.S. origin).

Table 6 also reports results for the impact of retailing channels, relative to the reference (grocery store). It seems the highest premia are associated with beverage stores and liquor stores (about 5%), and a similar price difference is also associated with the "all other" distribution channel, whereas the largest discount is provided by warehouse clubs.

B. Hedonic model with interaction effects

The wine literature recognizes the existence of possibly important variety-by-location interactions (Alston, Anderson, and Sambucci, 2015). Therefore, in trying to separate the roles of varietals and geographic origin in affecting prices, it may be desirable to account for such interactions. In addition to the baseline hedonic model discussed in the foregoing, we also estimate a model that includes interaction terms between geographic origin and varietal indicator variables. In other words, we expand the model of Equation (1) to include a set of interaction terms $z_g z_v$, where z_g denotes geographic origin indicator variables and z_v denotes varietal indicator variables. Because we have 38 varietal indicator variables and 81 geographic origin indicator variables, in principle, this adds 3,078 additional explanatory variables to the hedonic price equation. However, many of these interaction effects have zero observations in the sample (e.g., no Barbera in Willamette Valley), such that the net addition of the interaction model amounts to 729 variables.

The results for the interaction model are reported in the last three columns of Table 6. The coefficients and the implicit prices reported there pertain to the total effect evaluated at the relevant conditional means. For example, if the kth attribute of interest is the AVA "Carneros," then the coefficient reported for the interaction model is $\beta_k + \sum_v \delta_{kv} E[z_v | AVA = k]$, where, as before, β_k is the coefficient of the stand-alone Carneros indicator variable, and δ_{kv} denotes the parameters of the interaction variables between the Carneros region and varietal indicators. The standard error for this total coefficient, and the associated p-value, are calculated using the delta method. Given such a calculation of the interaction total coefficient, the implicit price, expressed as a percent, is then computed in the same fashion as noted earlier.

From Table 6, we see that the model with interaction effects entails a large price discount for white wine relative to the reference of red wine (–7.1%). As for the implicit prices of varietals and geographic origin attributes, the results of the interaction model are quite similar to those of the baseline model. For example, the correlation coefficient between the implicit prices of varietals across the two models is 0.924. Some minor differences, resulting from introducing interaction effects, include a larger discount for Cabernet blends and Riesling, and a larger premium for Zinfandel and Chardonnay. As for the implicit prices for geographic origin attributes, they are extremely close: the correlation coefficient between the implicit prices of the baseline and the interaction models is 0.995. Among the few notable differences that we observe is that the model with interaction effects predicts larger premia for Napa wines (e.g., the Napa Valley premium increases from 28.4 to 44.8%).

We do not necessarily believe that the interaction model is "better" for the purposes of the research question in this paper. At a minimum, however, this provides a robustness check on the conclusions we derive vis-à-vis the implicit prices of varietals and geographic origin. The conclusion from the results in Table 6 is that the two models are quite consistent, and both indicate moderate impacts of varietals and larger impacts of geographic origin on estimated implicit prices. We note, again, that these are ceteris paribus effects, after accounting for other likely sources of product differentiation, including extensive controls for wine brands.

C. Relative importance of attributes: Shapely values

Having established that the explanatory variables included in the hedonic price equations have non-negligible price impacts, as evidenced by the computed implicit prices, here we investigate the relative importance of the various attributes in the determination of equilibrium wine prices. In particular, just how important is the geographic origin information provided by AVA, county, and state appellations? How does geography as a marketing tool compare with, say, the role of variety information? To answer these questions, we apply a technique that assesses the relative importance (influence) that independent variables contribute to a model's predictive abilities. The approach is inspired by the Shapley value concept from cooperative game theory, which is finding widespread interest in modern machine learning methods (Lundberg and Lee, 2017).

Our application of the Shapley value approach follows the original development provided by Lipovetsky and Conklin (2001) to evaluate the relative importance of individual variables in a linear model's prediction. The main criterion to assess the latter is the regression fit as measured by the R^2 statistics. The key is a decomposition of this metric that provides attribution shares to each right-hand-side variable. To see how this objective is related to one of the more celebrated concepts in game theory, it is helpful to recall the basics of the Shapley value (see Roth (1988) for a lucid discussion). The setting is one where any subset of the set $N \equiv \{1, 2, \ldots, n\}$ of players can create value by cooperating. The latter is measured by the characteristic function v that maps any coalition of a subset $S \subseteq N$ of players to a real number $v(S)$ that summarizes the overall value of the game. How much does each individual participant contribute to the coalition? One could start by looking at the marginal contribution of player i when participating in a coalition S, that is, $\Delta_i = v(S) - v(S - \{i\})$. The problem, of course, is that this marginal value depends on the specific subset S one considers. Shapley (1953) provides an axiomatic formulation where, nonetheless, a unique "Shapley value" $\phi_i(v)$ for each player can be obtained such that the value of the grand coalition $v(N)$ is fully distributed, that is, $\sum_{i \in N} \phi_i(v) = v(N)$.

The Shapley value is, essentially, the average over all possible marginal contributions Δ_i (i.e., the marginal contribution of player i evaluated with respect to all possible coalitions it can be part of). That is, the Shapley value can be expressed as

$$\phi_i(v) = \sum_{\forall S \subseteq N} w(S)[v(S) - v(S - \{i\})], \tag{2}$$

where the weights associated with each coalition are $w(S) = (m-1)!(n-m)!/n!$, with m denoting the size of coalition S (which, as noted earlier, n is the number of all possible players, that is, the size of the grand coalition).[10]

The parallel with the problem at hand is apparent as we now conceive of individual explanatory variables as "players" that cooperate in a regression, where the value of their cooperation is measured by the quality of the model's fit. As a metric for the latter, Lipovetsky and Conklin (2001) focus on the traditional R^2 statistics. The beauty and apparent notional simplicity of the Shapley value in Equation (2) hides challenging computational issues, however. With a set of n players, one needs to evaluate the coalition value $v(\cdot)$ for $(2^n - 1)$ possible coalitions—a magnitude that increases exponentially with the number of players n. In our context, for a regression involving n explanatory variables, to impute a Shapley value contribution to each individual regressor, one needs to run $(2^n - 1)$ regressions, something that is challenging for even moderate regression sizes, and clearly not feasible in our context.[11]

The economic question of interest, however, does not concern individual regressors but rather the value contribution of sets of regressors. That is, we are not so much interested in how the information concerning one AVA adds much to explaining wine equilibrium prices; we care more about understanding whether the set of all

[10]Beyond its central role in cooperative game theory, the Shapley value has been utilized in numerous applications in many disciplines. Of interest to the readers of this *Journal*, Ginsburgh and Zang (2012) show how the Shapley value can be used to rank wines based on the subjective assessment of a set of judges.

[11]For example, for a model with $n = 10$ regressors, the number of regressions is approximately one thousand; for $n = 30$ regressors, the number of regressions exceeds one trillion. In our model, including all fixed effects, we have approximately 5,000 right-hand-side variables.

AVAs, collectively, adds in a meaningful way to explaining equilibrium wine prices. Thus, to assess such broad features of the empirical hedonic price model, we implement the Shapley value decomposition of the model's R^2 for groups of variables rather than for individual regressors. For Shapely value imputation, we specifically consider six groups, which include: (i) wine type, (ii) varietals, (iii) geographic origin, (iv) market characteristics (e.g., retail channel type), (v) brand fixed effects, and (vi) year, season, and region fixed effects.

Given the current specification, we have 63 "coalitions" of explanatory variables. Let subscripts a, b, c, d, e, and f denote each of these six groups of regressors. Then the grand coalition, including all explanatory variables, will have a model fit that is denoted R^2_{abcdef}. Omitting all variables except variety-related variables, on the other hand, would lead to a model with fit denotes R^2_b, while omitting varietals but including everything else yields a model fit R^2_{acdef}. Hence, the marginal value of feature b, relative to the empty set, is simply R_b (recall that $v(\emptyset) = 0$). The marginal value of feature b when all other features are present, on the other hand, is $R^2_{abcdef} - R^2_{acdef}$. In this fashion, we can construct the marginal value of feature b relative to all possible coalitions, such that we can implement Equation (2) and construct the Shapley value for this feature. And this can be done for all of the six features of interest that we have identified.[12] By construction, these Shapley values will satisfy $\phi_a + \phi_b + \phi_c + \phi_d + \phi_e + \phi_f = R^2_{abcdef}$.

The results for the Shapley value decomposition are reported in Table 7. Using the R^2 metric of the baseline model, we find that, for equilibrium prices, brand-fixed effects are the most important determinants, accounting for 72%. Geographic origin is next, accounting for 10.7%, whereas varietals account for 6.2% and wine type for 1.4%. The two sets of control variables that we include in the hedonic price equation (e.g., retail distribution channels and other fixed effects) account for about 10%.

The large relative importance that Shapley values associate with brands vindicates the presumption, noted in the introduction, that repeat purchases are important in markets with experience goods, such that the provision of quality can be supported by credible branding strategies. Notwithstanding the role played by firms' individual brands in the wine market, however, it is apparent that "collective" branding messages conveyed by geographic origin are quite important. This is an accepted fact for Old World wine marketing (e.g., Castriota, 2020), but the foregoing results imply that, even for New World wines, credible certification of geographic origin matters. In fact, we find that the relative importance of geographic origin clearly exceeds the role played by varietals. This is consistent with other work showing that region of origin is usually more significant as a determinant of price than grape variety (e.g., Steiner (2004) for Australia, and Costanigro, McCluskey and Mittelhammer (2007) for North America).

The decomposition in the first two columns of Table 7 pertains to the R^2 for Equation (1), which fits the log of price. From an economic point of view, of course, the implied fitted price from the estimated equation is what matters most. From the estimated semilog model, with log fitted values $\hat{y}_i \equiv \ln p_i - \hat{\varepsilon}_i$, with a normal distribution the implied fitted price is $\hat{p}_i = \exp(\hat{y}_i) * \exp(\hat{\sigma}^2/2)$, where $\hat{\sigma}^2$ is the estimated variance of the regression error. Implied price residuals are thus $\hat{u}_i \equiv p_i - \hat{p}_i$,

[12]See Lipovetsky and Conklin (2001) for additional details on the construction of the required algorithm.

Table 7. Shapley values

	Model R^2		Implied $\tilde{R}^2$	
Baseline hedonic model	Shapley	%	Shapley	%
Wine type	0.011	1.35	0.012	1.65
Varietals	0.050	6.18	0.032	4.35
Geographic origin	0.086	10.72	0.088	11.98
Brand fixed effects	0.577	72.02	0.557	75.41
State control, excise tax, retail density, channel type	0.048	6.01	0.033	4.50
Year, quarter, and market fixed effects	0.030	3.73	0.016	2.12
Total (R2)	0.802	100	0.738	100
Hedonic model with Interactions				
Wine type	0.011	1.36	0.012	1.58
Varietals	0.060	7.44	0.050	6.59
Geographic origin	0.097	11.98	0.106	13.97
Brand fixed effects	0.565	69.75	0.543	71.54
State control, excise tax, retail density, channel type	0.047	5.80	0.033	4.35
Year, quarter, and market fixed effects	0.030	3.70	0.015	1.97
Total (R2)	0.810	100	0.759	100

Notes: This table reports the Shapley value metrics for the relative importance of sets of variables.

and the implied coefficient of determination for predicting prices themselves is $\tilde{R}^2 = 1 - \sum_i \hat{u}_i^2 / \sum_i (p_i - \bar{p})^2$. The Shapley value decomposition based on $\tilde{R}^2$ is also reported for the baseline model in Table 7. The results are similar to those based on the semilog R^2, and in fact, further emphasize the main take-home points: brands are the most important determinants, geographic origin is next, and geographic origin matters quite a bit more than varietals.

The second part of Table 7 reports the results for the model with interaction variables. The results are broadly consistent with those of the baseline model. The main impact of adding interaction effects, relative to the baseline, is to moderate slightly the impact of firms branding, which decrease from 75.4 to 71.5%, and to increase the role of geographic origin, which increases from 12 to 14% (for the metric based on $\tilde{R}^2$).

VI. Conclusion

In this paper, we analyze the determinants of retail wine prices in the U.S. market for wine consumed at home, with the objective of characterizing the main dimensions of product differentiation. We focus on U.S.-produced wines, sold in standard 750 ml bottles, and rely on an extensive dataset from Nielsen homescan data, obtained from a large and representative sample of household purchases from 2007 to 2019. Data extracted from products' UPC label description permits us to construct a large set of controls, including individual products' brands, wine varietal, and

https://doi.org/10.1017/jwe.2022.23 Published online by Cambridge University Press

geographic origin. Our hedonic price model also includes extensive controls for other local factors likely affecting retail prices.

Our empirical findings suggest that, after accounting for firms' brands, wine types, varietals, and other control variables, information about the geographic origin of U.S. wines carries considerable explanatory power. In particular, AVA labels are associated with significant price premia, relative to an undefined U.S. origin, over and above that secured by the effects of firms' brands. Our paper also proposes the use of an attractive metric to characterize the overall importance of a set of determinants of wine price, a measure based on the Shapley value from cooperative game theory. Using this metric, we find that over 70% of U.S. retail wine prices are accounted for by individual wines' own brands, a finding consistent with the basic economics of markets for experience goods and the role of reputation. We also find that geographic origin is the next most important predictor of wine prices, again even after accounting for the effects of brands. In particular, Shapley values suggest that the contribution of geographic origin is considerably greater (about twice as large) than that of varietals.

These findings have interesting implications for wine producers' marketing strategies and for policy. Whereas it is accepted that geographic origin information is essential for marketing Old World wines, especially wines from the European Union, conventional wisdom holds that such factors matter less for New World wines. Indeed, one of the marketing tools emphasized by New World wineries has been the use of varietals. Notwithstanding that, however, starting in the 1980s, the U.S. industry developed a standardized system of geographic origin denominations, centered on AVAs, that provides collective labeling options similar to those widely used in Europe. Our results vindicate this evolution and suggest that, indeed, AVA labeling is an important element of product differentiation, highly complementary to firms' marketing strategies captured by their own brands.

Acknowledgments. The authors thank the editor, Karl Storchmann, and an anonymous reviewer for their constructive comments.

Researcher(s)' own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researcher(s) and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

The author(s) declare that they have no competing interests.

References

Akerlof, G. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, 84(3), 488–500.

Alston, J. M., Anderson, K., and Sambucci, O. (2015). Drifting towards Bordeaux? The evolving varietal emphasis of US wine regions. *Journal of Wine Economics*, 10(3), 349–378.

Bajari, P., and Benkard, C. L. (2005). Demand estimation with heterogeneous consumers and unobserved product characteristics: A hedonic approach. *Journal of Political Economy*, 113(6), 1239–1276.

Bartik, T. J. (1987). The estimation of demand parameters in hedonic price models. *Journal of Political Economy*, 95(1), 81–88.

Box, G. E., and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2), 211–243.

Brown, J. N., and Rosen, H. S. (1982). On the estimation of structural hedonic price models. *Econometrica*, 50(3), 765–768.

Cardebat, J. M., and Figuet, J. M. (2004). What explains Bordeaux wine prices? *Applied Economics Letters*, 11(5), 293–296.
Carew, R., Florkowski, W. J., and Meng, T. (2017). Segmenting wine markets with diverse price functions: Evidence from California red and white wines sold in British Columbia. *Wine Economics and Policy*, 6(1), 48–59.
Castriota, S. (2020). *Wine Economics*. Cambridge, MA: MIT Press.
Combris, P., Lecocq, S., and Visser, M. (1997). Estimation of a hedonic price equation for Bordeaux wine: Does quality matter? *Economic Journal*, 107(441), 390–402.
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.
Costanigro, M., McCluskey, J. J., and Mittelhammer, R. C. (2007). Segmenting the wine market based on price: Hedonic regression when different prices mean different products. *Journal of Agricultural Economics*, 58(3), 454–466.
Ekeland, I., Heckman, J. J., and Nesheim, L. (2002). Identifying hedonic models. *American Economic Review*, 92(2), 304–309.
Epple, D. (1987). Hedonic prices and implicit markets: Estimating demand and supply functions for differentiated products. *Journal of Political Economy*, 95(1), 59–80.
Gandhi, A., and Nevo, A. (2021). Empirical models of demand and supply in differentiated products industries. In K. Ho, A. Hortacsu, and A. Lizzeri (eds.), *Handbook of Industrial Organization*, vol. 4, 63–139. Amsterdam, Netherlands: Elsevier.
Ginsburgh, V., and Zang, I. (2012). Shapley ranking of wines. *Journal of Wine Economics*, 7(2), 169–180.
Gorman, W. M. (1956). The demand for related goods: A possible procedure for analysing quality differentials in the egg market. *Iowa Agricultural Experiment Station*, Journal Paper No. 2319, November. [Reprinted in: *Review of Economic Studies*, 47, June 1980, 843–856.]
Gustafson, C. R. (2011). *Experimental Economics and Hedonic Pricing: An Application to Wine Attributes*. Davis, CA: University of California, Davis.
Gustafson, C. R., Lybbert, T. J., and Sumner, D. A. (2016). Consumer knowledge affects valuation of product attributes: Experimental results for wine. *Journal of Behavioral and Experimental Economics*, 65, 85–94.
Halvorsen, R., and Palmquist, R. (1980). The interpretation of dummy variables in semilogarithmic equations. *American Economic Review*, 70(3), 474–475.
Horowitz, I., and Lockshin, L. (2002). What price quality? An investigation into the prediction of wine-quality ratings. *Journal of Wine Research*, 13(1), 7–22.
Kwon, O. S., Lee, H., and Sumner, D. A. (2008). Appellation, variety, and the price of California wines. *ARE Update*, 11(4), 15–19.
Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy*, 74(2), 132–157.
Landon, S., and Smith, C. E. (1997). The use of quality and reputation indicators by consumers: The case of Bordeaux wine. *Journal of Consumer Policy*, 20(3), 289–323.
Landon, S., and Smith, C. E. (1998). Quality expectations, reputations and price. *Southern Economic Journal*, 64(3), 628–47.
Levaggi, R., and Brentari, E. (2014). The hedonic price for Italian red wine: Do chemical and sensory characteristics matter? *Agribusiness*, 30(4), 385–397.
Lipovetsky, S., and Conklin, M. (2001). Analysis of regression in game theory approach. *Applied Stochastic Models in Business and Industry*, 17(4), 319–330.
Lockshin, L., and Corsi, A. M. (2012). Consumer behaviour for wine 2.0: A review since 2003 and future directions. *Wine Economics and Policy*, 1(1), 2–23.
Lockshin, L., Jarvis, W., d'Hauteville, F., and Perrouty, J. P. (2006). Using simulations from discrete choice experiments to measure consumer sensitivity to brand, region, price, and awards in wine choice. *Food Quality and Preference*, 17(3–4), 166–178.
Lundberg, S. M., and Lee, S. (2017). A unified approach to interpreting model predictions. In U. von Luxburg and I. Guyon (eds.), *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 4768–4777. Red Hook, NY: Curran Associates Inc.
Meloni, G., and Swinnen, J. (2013). The political economy of European wine regulations. *Journal of Wine Economics*, 8(3), 244–284.
Menapace, L., and Moschini, G. (2012). Quality certification by geographical indications, trademarks and firm reputation. *European Review of Agricultural Economics*, 39(4), 539–566.

Mérel, P., Ortiz-Bobea, A., and Paroissien, E. (2021). How big is the "lemons" problem? Historical evidence from french wines. *European Economic Review* 138, https://doi.org/10.1016/j.euroecorev.2021.103824.

Nesheim, L. (2008). "Hedonic prices." In Palgrave Macmillan (eds.), *The New Palgrave Dictionary of Economics*. London: Palgrave Macmillan, doi.org/10.1057/978-1-349-95121-5_2370-1.

Oczkowski, E. (1994). A hedonic price function for Australian premium table wine. *Australian Journal of Agricultural Economics*, 38(1), 93–110.

Oczkowski, E., and Doucouliagos, H. (2015). Wine prices and quality ratings: A meta-regression analysis. *American Journal of Agricultural Economics*, 97(1), 103–121.

Orth, U. R., Lockshin, L., and d'Hauteville, F. (2007). The global wine business as a research field. *International Journal of Wine Business Research*, 19(1), 5–13.

Outreville, J. F., and Le Fur, E. (2020). Hedonic price functions and wine price determinants: A review of empirical research. *Journal of Agricultural & Food Industrial Organization*, 18(2). Available at https://www.researchgate.net/publication/338535129_Hedonic_Price_Functions_and_Wine_Price_Determinants_A_Review_of_Empirical_Research.

Pakes, A. (2003). A reconsideration of hedonic price indexes with an application to PC's. *American Economic Review*, 93(5), 1578–1596.

Roma, P., Di Martino, G., and Perrone, G. (2013). What to show on the wine labels: A hedonic analysis of price drivers of Sicilian wines. *Applied Economics*, 45(19), 2765–2778.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, 82(1), 34–55.

Roth, A. E. (1988). Introduction to the Shapley value. In A. E. Roth (ed.), *The Shapley Value: Essays in Honor of Lloyd S. Shapley*, 1–27. New York: Cambridge University Press.

Santiago, M., and Sykuta, M. (2016). Regulation and contract choice in the distribution of wine. *Journal of Wine Economics*, 11(2), 216–232.

Schamel, G. (2002). California wine winners: A hedonic analysis of regional and winery reputation indicators. Paper presented at the American Agricultural Economics Association Annual Meeting, Long Beach, CA. Available at https://www.researchgate.net/publication/23504920_California_wine_winners_a_hedonic_analysis_of_regional_and_winery_reputation_indicators.

Schamel, G. (2006). Geography versus brands in a global wine market. *Agribusiness: An International Journal*, 22(3), 363–374.

Schamel, G. (2009). Dynamic analysis of brand and regional reputation: The case of wine. *Journal of Wine Economics*, 4(1), 62–80.

Schamel, G., and Anderson, K. (2003). Wine quality and varietal, regional and winery reputations: Hedonic prices for Australia and New Zealand. *Economic Record*, 79(246), 357–369.

Shapiro, C. (1983). Premiums for high quality products as returns to reputations. *Quarterly Journal of Economics*, 98(4), 659–679.

Shapley, L. S. (1953). A value for n-person games. In H. W. Kuhn and A. W. Tucker (eds.), *Contributions to the Theory of Games*, vol. II, 307–17. Princeton, NJ: Princeton University Press.

Steiner, B. E. (2004). Australian wines in the British wine market: A hedonic price analysis. *Agribusiness: An International Journal*, 20(3), 287–307.

Storchmann, K. (2012). Wine economics. *Journal of Wine Economics*, 7(1), 1–33.

Taber, G. M. (2005). *Judgment of Paris, California vs. France and the Historic 1976 Paris Tasting That Revolutionized Wine*. New York: Scribner.

Tirole, J. (1988). *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.

Troncoso, J. L., and Aguirre, M. (2006). Price determinants of Chilean wines in the US market: A hedonic approach. *Spanish Journal of Agricultural Research*, 4(2), 124–129.

Young, D. J., and Bielińska-Kwapisz, A. (2002). Alcohol taxes and beverage prices. *National Tax Journal*, LV(1), 57–73.

https://doi.org/10.1017/jwe.2022.23 Published online by Cambridge University Press

Appendix A

Table A1. Wine products in the Nielsen consumer panel data

			Percentiles				
Module	% Share (quantity)	Mean	1st	5th	50th	95th	99th
Aperitifs	0.01	16.7	3.2	9.9	17.8	23.7	27.9
Domestic dry table	64.66	6.7	1.6	2.1	5.5	15.1	22.7
Flavored	4.86	5.4	2.0	2.7	4.2	12.0	16.5
Imported dry table	17.99	8.1	2.3	3.2	7.1	15.5	22.2
Kosher table	0.69	5.9	2.9	3.5	4.9	13.2	17.0
Non-alcoholic	3.74	3.08	1.16	1.78	2.91	5.46	7.99
Sake	0.19	11.0	2.5	3.4	8.3	24.5	36.5
Sangria	2.76	4.3	1.5	1.8	3.7	8.3	12.2
Sparkling	3.70	11.4	3.5	4.4	9.4	22.3	50.1
Sweet dessert domestic	0.70	6.5	3.1	4.0	6.0	9.2	19.7
Sweet dessert imported	0.19	18.0	5.7	8.3	14.8	35.0	95.6
Vermouth	0.51	6.6	2.9	3.6	5.9	11.4	17.3
Total	100						
Distribution by packaging							
Standard 750 ml bottle	42.74	8.43	2.09	2.71	7.55	16.69	25.78
bulk (> 750 ml)	53.41	4.10	1.48	1.83	3.99	6.83	8.62
small (<750 ml)	3.85	7.14	1.86	2.56	5.69	17.54	37.24

Notes: Summary statistics are weighted using the Nielsen projection factor. Prices are expressed as $ per 750 ml volume (bottle equivalent). Distribution by packaging is across data pooled from all 12 modules.
Source: Nielsen Consumer Panel Data.

Table A2. List of top 50 brands in terms of market share (names if alphabetical order)

14 Hands	Duplin	Mirassou
Apothic Red	Estancia	Oak Creek
Barefoot	Fetzer	Oak Leaf Vineyards
Bay Bridge Vineyards	Flip Flop	Oliver
Beaulieu Vineyard	Fox Brook	Ravenswood
Beringer	Francis Coppola	Redwood Creek
Blackstone	Gallo Family Vineyards	Robert Mondavi
Bogle	Gnarly Head	Rodney Strong
Challis Lane	H R M Rex-Goliath	Smoking Loon
Charles Shaw	Hogue	Sterling Vintner's Collection
Chateau Ste Michelle	J. Lohr Estates	Sutter Home
Clos Du Bois	Josh	The Naked Grape
Columbia Crest	Kendall-Jackson	Three Wishes
Crane Lake	La Crema	Tisdale
Control Brand	Mark West	Turning Leaf
Cupcake Vineyards	Ménage à Trois	Winking Owl
Dark Horse	Meridian	

Notes: This table provides the list of the top 50 brands (in terms of quantity). Names are listed in alphabetical order and provided as mentioned in the data.
Data Source: Nielsen Consumer Panel Data (domestic dry table wine module). These top 50 brands account for about 68% of the total quantity. The hedonic price equation also includes dummies for approximately 3,939 other brands that account for the remaining 32% of quantity in the market.

Table A3. Box Cox regression

	θ	Log likelihood
Box Cox	0.07 (.0013)	−2,597,600
Linear	1	−2,905,126
Semi-Log	0	−2,598,927

Notes: This table reports the result of a Box-Cox regression where the left-hand-side of Equation (1) is represented as $(p^\theta - 1)/\theta$. Thus, this specification nests the semi-log functional form used in the text, which attains when $\theta \to 0$, as well as the simple linear model when $\theta = 1$ (Box and Cox, 1964). Because of the large number of variables in our model (about 4,000), and the large number of observations (about one million), estimation of the Box-Cox model proved exceedingly burdensome computationally. The Box-Cox parameter in this table, obtained using the Stata procedure "boxcox," uses all the right-hand-side of Equation (1) except brand fixed effects.
Data Source: Nielsen Consumer Panel Data.

Appendix B: Extracting text data from UPC description

We extract the wine attribute information, such as type, varietal, and geographic origin, from the UPC description in the Nielsen data. Table B1 presents some examples of UPC and brand descriptions from the Nielsen data. Varietal and geographic label columns show information extracted from UPC descriptions. We used regular expressions in Stata to code this information extraction. For most UPC descriptions, the first word is generally the band name. Brand information is also provided by Nielsen as a separate variable. The remaining letters stand for country name, region, and varietal. For instance, CHRD is Chardonnay, P-GR is Pinot Grigio, MLBC is Malbec, MRLT is Merlot, WT is White, DDT is Domestic Dry Table, BLS is Blush, etc. We did find some redundancies, where more than one abbreviation for one word could be used (e.g., Moscato could be MSC, MSCT, or MUSCATO). We did our best to cross-check such entries. Further, to ensure that these abbreviations mean what we take them to mean, we cross-checked our inferences by matching brand names and UPC descriptions with information at the site www.wine-searcher.com, which has a comprehensive list of wines with brand names, grape blends, and geographic origins.

Wine brand names also come with small variations, and Nielsen assigns new codes to the brands even with the slightest variation in brand description. In this study, we are treating different variants of a brand as one brand. For instance, Francis Coppola dmnd cltn, Francis Coppola diamond cllctn, Francis coppola presents, Francis Coppola Director's cut are considered as one brand **Francis Coppola**. As another example, brand **Robert Mondavi** includes: Robert Mondavi, Woodbridge Rbrt Mndv, Robert Mondavi Private Selection, La Famiglia di Robert Mondavi, Woodbridge Rbrt Mndv Slt Vy Sr, and Woodbridge by Robert Mondavi.

Table B1. Examples of UPC description from Nielsen data

Brand description	UPC description	Varietal	Geographic label
LAKE SONOMA WINERY	LSW P-N SCS V RED DDT	Pinot Noir	Sonoma Coast
ANGLE'S SECRET	ANG-ST P-SRH HG VLY V RED DDT	Petite Syrah	High Valley
BERINGER	BGR WT MRLT NV CA V BLS DDT	Merlot	Napa Valley
BLACK BOX	BK-B CB-S PS-R V RED BB DDT	Cabernet Sauvignon	Paso Robles
MCMANIS	MCMANIS BARB CA V RED DDT	Barbera	California
MATTHEW FOX	MATTHEW FOX CB-S CA V RED DDT	Cabernet Sauvignon	California
DOMAIN ST GEORGE	DM-SG WT ZN V BLS DDT	Zinfandel Rose	United States
CASTLE ROCK	C-R P-GRS WV V WT DDT	Pinot Grigio	Willamette Valley
CONSILIENCE	CONSILIENCE SYR SBC V RED DDT	Syrah	Santa Barbara County
FREY	FREY GWRZT OR MDCN V WT DDT	Gewurztraminer	Mendocino

Cite this article: Chandra, R., and Moschini, G. (2022). Product differentiation and the relative importance of wine attributes: U.S. retail prices. *Journal of Wine Economics*, 17(3), 177–208. https://doi.org/10.1017/jwe.2022.23