AMERICAN ASSOCIATION OF WINE ECONOMISTS

AAWE WORKING PAPER
No. 229
Economics

UNCORKING EXPERT REVIEWS WITH SOCIAL MEDIA: A CASE STUDY SERVED WITH WINE

Alex Albright, Peter Pedroni, and Stephen Sheppard

Jun 2018

www.wine-economics.org
Uncorking Expert Reviews with Social Media:
A Case Study Served with Wine*

Alex Albright†, Peter Pedroni‡, Stephen Sheppard§
June 5, 2018

Click here for most recent version.

Abstract

The growth of social media outlets in which individuals post opinions on publicly consumed goods provides an interesting and relatively unexplored area for examination of the role of crowd sourcing amateur opinions in areas traditionally relegated to experts. In this paper we use wine as an illustrative example to investigate the interaction between social media and expert reviews in the market for high end consumer goods. In particular, we exploit a novel data set constructed from the social media website CellarTracker, which is composed of the averaged individual reviews for 355 distinct wines on a quarterly basis from 2004 through 2017, and pair this with a similarly dimensioned panel of average auction prices for these wines as well as the reviews from three leading experts. We develop a signal extraction model to motivate the interaction between amateurs and experts in revealing a measure of the quality of the wine. The model is then used to motivate the adaptation of an empirical panel structural VAR approach based on Pedroni (2013) by embedding the expert reviews as an event analysis within the panel VAR, which is used to decompose information into components that signal the quality of the liquid in the bottle versus other aspects of the wine that are valued by the market. The approach also allows us to decompose the influence of the expert reviews into components associated with what we define as the quality of the wine versus the pure reputation effect of the expert. The results on expert reviews are consistent with the idea that experts can substantially impact prices through channels other than their signals of quality.

*We are immensely grateful to Peter Gibson and Eric LeVine for significant help with the data collection process. We thank the American Association of Wine Economists and Harvard for support of this research through a Research Scholarship and a Summer Research & Travel Grant, respectively. We are grateful to Jon Bakija, Peter Montiel, Quamrul Ashraf, and Richard De Veaux for thoughtful discussions and feedback, Julien Miquel, Jonathan Clark, Conor Dowling, and Andrea Danyluk for their assistance in early stages of data collection, conference participants at the 2016 American Association of Wine Economists Conference and the 2015 Connected Life Conference for helpful comments, and Colin Williams for research assistance.

†Harvard University. Email: apalbright@g.harvard.edu
‡Williams College. Email: ppedroni@williams.edu
§Williams College. Email: sshedpar@williams.edu
1 Introduction

The impact of expert reviews on market price has been a perennial favorite subject of study for high-end consumer goods. With such studies, there is often a temptation to attribute the impact of an expert review entirely to the reputation of the reviewer (e.g., Reinstein and Snyder, 2005; Ali and Nauges, 2007). Depending on one’s view of the nature of expert reviews, this can lead to the impression that market price in the world of so-called luxury goods is highly influenced by the proclamations of a handful of expert reviewers.

Alternatively, many markets now have extensive mechanisms for recording and transmitting amateur reviews about product quality. These mechanisms have been a core feature of vendors arranging for or advertising the sale of used goods (eBay), services (OpenTable or Airbnb), or other differentiated products (Amazon, Rotten Tomatoes). The past 10-15 years has seen the growth of similar mechanisms for sharing information about the differentiated luxury goods that had been the exclusive province of expert reviewers.

The simultaneous existence of amateur and expert reviews about the same product raises the interesting possibility that it might be possible to use amateur reviews, where the reputation of the reviewer is unknown or has minimal influence, to untangle two components that are combined in expert reviews: the quality of the product under review and the reputation of the reviewer. In this paper, we argue that the information from expert reviews constitutes a component of a broader signal extraction problem undertaken by consumers of differentiated luxury goods, and that it is important not to conflate the quality signal component of the expert reviewer with the purely reputational effect of the reviewer, which is independent of the quality signal.

To empirically address the market for high-end goods, we hone in on an example that is so steeped in culture that its appreciation is often intimidating to the average individual—that is, wine. The market for high-end wines represent a leading examples of the phenomenon of potentially large influences from expert reviews.1 In a dichotomy reminiscent of an open question in Ashenfelter and Jones (2013)’s conclusion, we address the mechanism behind such effects, i.e. the extent to which experts influence the market through the signal they provide regarding quality information versus other channels independent of quality information. In particular, the recent growth of the social media forum CellarTracker.com provides an interesting opportunity to track a dimension of the consumer evaluation of wines that can be used to help interpret price responses to expert opinions. Toward this end, we have constructed a unique large-scale time series panel of amateur wine reviews obtained from CellarTracker, which we pair with similarly dimensioned panels of auction prices for the corresponding wines, and supplement with extensive data on expert reviews.

The aspects of complex, unknown, interdependent dynamics that underlie the evolution of wine quality and wine reviewing fall squarely in the realm of what structural vector autoregression (VAR) analysis is intended to address in the time series literature. In this context, structural VAR analysis can aid in more nuanced decomposition of the quality signal from expert reviews and their impact on the market. We therefore use our dataset in conjunction with the recently developed panel structural vector autoregressive methods of Pedroni (2013) to decompose noisy signals into their quality components by using restrictions motivated by a simple signal extraction model. This enables us to furthermore embed the expert reviews in the panel VAR analysis and decompose the effects of the reviews of well-known wine expert organizations, those of Robert Parker, Jancis Robinson, and Antonio Galloni, into their component parts – quality information and other components. Our results on the experts are consistent with the idea that expert reviews in high-end markets can meaningfully move prices via associated publicity effects alone.

The remainder of this paper is structured as follows. Section 2 describes the unique and fitting nature of wine as a case study. Section 3 discusses related literature and concepts. Section 4 describes the data used in our analysis. Section 5 presents some initial reduced-form evidence on the influence of experts on prices in the high-end wine market. Section 6 discusses a possible mechanism behind such effects and introduces how we think of the signal extraction problem. Section 7 describes how we implement the panel structural VAR approach. Section 8 discusses our results and Section 9 concludes.

---

1 The market for high end wines is also a surprisingly large market in terms of dollar value. To give an example, one of the leading wine auction companies, Acker Merrall & Condit, recently announced plans to surpass $1 billion dollars in auction sales in June 2018, as noted in Acker, Merrall & Condit (2018).
2 The Appeal of Wine as a Case Study

Wine has a powerful connection to the world of the sophisticated that far surpasses that of any other food or drink. It is this tie to the sophisticated, and the social pressure that comes with it, that renders wine a great case study of the forces of quality and reputation. Perceptions are well-known to factor into wine quality judgments. In fact, color and labelings have been shown to successfully deceive even self-proclaimed experts (Morrot, Brochet and Dubourdieu, 2001; Lehrer, 2012). The susceptible nature of wine quality determination to suggestion makes it an apt context in which to study the impact of experts. Producers readily display experts’ scores on the shelves by specific wines in order to assure consumers of the accurate monetary worth of the wine. Since consumers have not experienced the wine ex ante, they are willing to take expert ratings as certified evidence they are not being sold a lemon.2

Expert reviews may sound similar to Consumer Reports (in the classic used cars market example) in their capacity to provide signaling and, thus, mitigate the Lemons problem that befalls markets with uninformed consumers (Ackerlof, 1970). However, wine possesses a unique series of characteristics that render it an apt case study for our purposes. For one, high-end wines do not merely represent leading examples of the phenomenon of expert influence on consumer evaluation, they also present rich and interesting signal extraction problems. In contrast to some luxury goods, the quality of a given bottle of wine is not fixed, but rather evolves over time. Moreover, the quality does not simply depreciate or appreciate over time, but rather tends to be non-monotonic, rising for a period before falling. Expert reviewers render opinions on quality not simply as a static notion, but in large part as a forecast of how the wine is expected to evolve. But, once it has been produced and released, the evolution of wine quality is exogenous. This feature makes wine attractive for the purposes of structural analysis because, unlike say restaurant reviews, the quality of a given bottle of wine is not responding endogenously to reviews.3 In other words, once a vintage is released, information is gradually accumulated regarding its quality, while the quality is also changing, but is (more or less) exogenous with respect to the reviews.4 In both car and wine markets, reviews are trying to forecast the quality of the product over time (as the car gains more miles, or as the wine ages), but the signal extraction problem for wine is more interesting in that, aside from perhaps very high end collectible automobiles, most automobiles simply decline monotonically in quality and value over time, while wine possesses the aforementioned nonlinear time profile, rising before declining.

Confounding the signal extraction problem, in the case of wine, is the fact that it is not altogether clear what expert reviewers intend as the comparison benchmark. Clearly a higher rating on a wine is not intended to convey that the wine is superior in quality to a lesser rated wine in some absolute sense, since the rating is presumably conditional on a number of unknown factors, including perhaps the category of wine, the vintage, the price of the wine, or perhaps even the reviewer’s prior expectation of the specific bottling of wine.5 The expert reviews, in turn, impact the consumer’s evaluation of the quality of wine as part of the signal extraction, and both, in turn, impact the evolution of market price. Also relevant to the signal extraction problem is the fact that supply for a given vintage changes over time as people make consumption decisions in response to accumulating information, while supply of a given model year of say an automobile more or less simply declines monotonically as cars depreciate and go out of service. For example, if information becomes available via social media or expert reviews, individuals may choose to consume a particular wine earlier rather than later, or vice versa. In short, both these examples highlight how the relationship between quality, evaluation of quality and price are intertwined in a dynamic and complex manner.

Recall informed consumers are necessary for price to be used as a signal of product quality by a monopolist, a château in the case of the wine market (Mahene, 2004). Also note that, as one would expect, there is empirical evidence that the higher the wine price, the more influential expert ratings (Gibbs, Tapia and Warzyński, 2009). This makes it maybe a bit more similar to reviews of automobiles, where reviews are often for a given model year (i.e. vintage). Recalls would be an exception. There is no cardinal measure for wines (Quandt, 2006).
3 Related Literature

This paper builds on related papers that also address the influence of experts on markets. Ashenfelter and Jones (2013) shows that expert ratings influence price beyond the extent to which they summarize the weather, a factor shown to predict mature wine prices (Ashenfelter, Ashmore and Lalonde, 1995). Our paper implicitly take up an open hypothesis from Ashenfelter and Jones (2013)'s conclusion – ‘it is also possible that the experts’ ratings influence prices because they create values that are independent of the function’ – by first presenting reduced-form evidence the effect of experts and then investigating the underlying mechanism behind such effects. To get under the hood of Parker effects specifically, we treat the topic as a signal extraction problem. This allows us to decompose many of the dynamics that are ingrained in previous work on herd behavior and consumer learning (e.g., Bikhchandani, Hirshleifer and Welch, 1998; Banerjee, 1992; Moscarini, Ottaviani and Smith, 1998) as well as work focusing on the relationship between reputation and quality (e.g., Shapiro, 1982; Smallwood and Conlisk, 1979; Hörner, 2002).

Network analyses provide ample theoretical research ground for understanding the dynamics behind judgment aggregation both with and without the presence of experts (e.g., Bozbay, Dietrich and Peters, 2014; Golub and Jackson, 2010; Elliott, Golub and Jackson, 2014). Particularly relevant is Golub and Jackson (2010), which shows that prominent groups of opinion leaders destroy the process of “efficient learning.” In short, the attention prominent groups receive cause their information to be overweighted and, thus, their idiosyncratic error leads the learning of the entire group astray. This area of research has become even ripe for empirical investigation given newly available swathes of individual-level consumer data, such as our wine dataset.

The recent growth of social media over the past decade has provided economists with interest in ratings and judgment aggregation on quality via many new sources. Academics have investigated review manipulation (Mayzlin, Dover and Chevalier, 2014), developed models for taste acquisition (McAuley and Leskovec, 2013), and evaluated the effect of aggregate judgments on restaurant revenues (Luca, 2016; Anderson and Magruder, 2012). High-frequency social media data often becomes synonymous with the popular science term, “the wisdom of the crowd.” Muchnik, Aral and Taylor (2013) uses a large-scale randomized experiment on a social news aggregation site to see whether ratings systems online can accurately “harness” the wisdom of the crowd to produce useful information about the product whose quality is being rated. The authors found that while negative social influence is corrected by the crowd, positive social influence is not, and can create ratings bubbles that would overstate a good’s quality.

In this paper, we discuss expert reviews as the reviews of individuals while we discuss the reviews of a community as the average among users in an online community. While there is empirical evidence that crowds have bested experts in their own predictions on objective measures, such as returns from the stock market (Nofer and Hinz, 2014), comparing the crowd and experts when it comes to luxury goods is not as straightforward. However, such results have spurred authors, such as Arora and Vermeylen (2013), to ask if end is near for “the expert” in the realm of art and other experience goods. With these ideas in mind, we use this article to investigate the mechanics of the effects of experts by harnessing our unique access to social media data, a form of crowd wisdom, in a high-end good market.

4 Data

We construct a unique time series panel combining auction prices, online reviews, and expert reviews. We collaborated with Peter Gibson of Wine Market Journal and Eric LeVine of CellarTracker to generate a unique dataset of 355 1990-2010 vintage wines that are rich in both auction price and amateur review data. We then manually collected expert reviews on the sample from Robert Parker’s RobertParker.com, Jancis Robinson’s JancisRobinson.com, and Anthonio Galloni’s Vinous.com.

4.1 Auction data

We use Wine Market Journal as our source for wine auction prices. The Wine Market Journal spans back to 1997 and tracks every solid-lot trade from all major European and U.S. auction houses. Internet auction data is also included since Wine Market Journal has data even beyond physical auction houses. We use only prices for 750ml bottle prices for the analysis in order to
remain consistent. We use auction data as our first filter for our wine sample since we care about conditioning on the wine being high-end. Specifically, with Peter Gibson’s help, we collect data on wines traded in at least 20 consecutive quarters since 2004, which yields a sample of 873 wines and 30,258 quarterly auction prices.

Since we want real auction prices, we deflate the nominal auction prices secured via Wine Market Journal using a monthly price series from the CPI called “at home alcoholic beverages” that covers our span of wine auction data (2004Q1-2017Q2).\footnote{We specify after 2004 since that is when CellarTracker was launched publicly and our goal is to combine these two data sources for wines.} Auction price discussions in forthcoming sections always refer to real auction prices.

4.2 Social Media Reviews
We collect informed consumer reviews from the social media website CellarTracker.com. Launched publicly in 2004, CellarTracker allows users to log wines that they own, and provide reviews (tasting notes) on those they have tried. The online community consists of half a million users as well as over 7 million tasting notes (as of Summer 2018). The network’s popularity within the wine community makes it a natural choice for empirically tracking the opinions of informed wine consumers. Furthermore, for high-end wines, the number of amateur reviews can be fairly large, and therefore difficult to manipulate by any one reviewer.\footnote{The “at home” terminology means excluding restaurant and bar prices, which is appropriate for our analysis. The base year for the series is 1983. But since we are always looking at auction prices in logs and looking at impulse responses for the changes, we are always looking at percent changes, so the base year should not matter much.}

Tasting notes on the site are time stamped and can consist of only a written review, only a point value, or, as is the case the majority of the time, a written review and a corresponding point value out of 100 that is determined by the reviewer.\footnote{The complete review history of each reviewer is easily available on CellarTracker. In principle, this allows one to easily discount reviews that were posted by individuals who have posted only a small number of anomalous reviews. (Although in practice, at least anecdotally, this type of manipulation does not appear to occur much.)} Our categorization of CellarTracker as an informed community is not made in a vacuum. McAuley and Leskovec (2013) use CellarTracker data to develop a recommendation system that accounts for the user’s level of experience. The authors develop models for notions of user evolution to discover acquired tastes. The results in the paper show that their model is most successful in the case of movie data and less successful for beer and wine. McAuley and Leskovec (2013) explain this is probably because there is a larger spectrum of expertise levels for movies, “whereas users who decide to participate on a beer-rating or wine-rating “website are likely to already be somewhat ‘expert.’”

With Eric LeVine’s help, we collected all tasting note data on the 873 wines we identified with adequate auction data. Since we want wines with rich auction and CellarTracker data, we require wines to have at least one CellarTracker review per quarter and at least 16 consecutive quarters that contain both tasting note scores and auction price averages. This restriction yields 355 wines with 10,109 quarters of auction price and CellarTracker data.

4.3 Expert Reviews
The status of the wine market shifted in the 1990s, when critics’ ratings became the new dominant factor in pricing. In fact, en primeur pricing (pricing when wine is still in barrels) came to almost entirely revolve around an American wine critic named Robert Parker, who began the newsletter The Wine Advocate in 1978. Parker’s ratings, which are also published with a verbal tasting note, use a 100-point scale, similar to CellarTracker, and have become among the most well known ratings to wine producers.\footnote{Here is an example tasting note from CellarTracker about a bottle of 1990 Haut-Brion: “A dark garnet color. The nose hints at classic bordeaux; tobacco, spice, cedar. The palate was surprisingly youthful for a near quarter century wine. The tannins are still present complemented with an acidity that belies its age. Notes of earth, tobacco, dark fruit, spice. No question, this is a classic wine that is not only drinking beautifully now, it has the ability to last many more years in the cellar. Gorgeous.” The numerical grade attached to this note is a 96.} Parker has been called the most influential of any critic in any field since he is the only critic in any field whose opinions are followed throughout the entire world (Barthélemy, 2010). His powerful reputation makes him an obvious choice for our investigation into the role of expert reviews.\footnote{Parker’s 100-point scale has elements of a cardinal measure of wine since the differences in grades seem to imply levels of differences in quality. However, Parker himself has said that it is possible that the difference between a 96, 97, 98, 99, or 100 could have been his emotion in the moment.}
Parker’s grades are available on RobertParker.com (with a subscription) and are accompanied by the month and year of the relevant tasting. For this analysis, we also collect expert reviews via JancisRobinson.com and Vinous.com, two other sites lead by prominent wine critics, Jancis Robinson, and Antonio Galloni. Their reviews are also available via subscription. We collected all Parker, Robinson, and Vinous reviews that occurred in the 10,109 quarters of price and review data – 240, 399, and 146 reviews, respectively – as well as the most recent Parker, Robinson, and Vinous reviews that predates the quarterly sample period for each wine. We therefore have data on expert review levels and changes.

4.4 Summary

Our dataset is an panel of 355 1990-2010 vintage wines, each of which has at least 16 consecutive quarters of both auction price and CellarTracker rating score averages. The dataset also includes three dimensions of expert review scores (Parker, Robinson, and Vinous). Our wine sample was created by setting requirements on richness of auction prices and social media restrictions. However, despite any regional restrictions, the sample is dominated by wines from Bordeaux (175) and California (109).

Figure 1: Histogram of Vintage Years

We offer a number of descriptive visuals to better describe our data. Figure 1 presents the frequencies of vintages 1990-2010 in our sample. Vintages 2000 and 2005 are those most represented (followed by 1990, 2001, and 2003). Meanwhile, vintages 1991, 1992, and 1993 are almost entirely missing in action. For each of the 355 wines, there is time series data over some span of consecutive quarters for mean real auction prices and mean CellarTracker scores. Figure 2 illustrates the two series for the 355 wines. Lastly, Figure 3 presents histograms of Parker, Robinson, and Vinous scores. The median scores are 95, 17.5, and 95 for Parker, Robinson, and Vinous, respectively. Clearly, Parker is more skewed towards the highest scores than are Robinson or Vinous. Robinson’s scores are the most balanced in distribution.

12 If there are multiple reviews from any reviewer in a quarter, we collect the average. If symbols such as + or - are included in the review, we ignore these and collect only the numerical component. We also collected the textual review components. These could be useful for further research.

13 Other regions include: Burgundy (1), Champagne (16), Piedmont (2), Rhone Valley (22), Tuscany (11), Campania (1), Castilla y Leon (1), South Australia (5), Washington (7).
Figure 2: Time Series Data

Time Series of Mean Real Auction Prices

Time Series of Mean CellarTracker Scores

Figure 3: Histogram of Expert Review Scores

RobertParker.com Scores

Information about the Robert Parker website

JancisRobinson.com Scores

Information about the Jancis Robinson website

Vinous.com Scores

Information about the Vinous website

via Albright et al. (2018)
5 Empirical Motivation

Our data constitutes a panel of 10,109 wine-quarter observations (355 wines, each over a minimum of 16 and maximum of 54 consecutive quarters). To motivate our eventual VAR approach, we first illustrate reduced-form estimates on how reviewer scores move prices. Specifically, we exploit variation in expert review timing and changes within wines (thanks to our in-depth expert review data collection) for a differences-in-differences approach.

We use a standard differences-in-differences specification that includes controls for both wine and quarter fixed effects. We investigate the impacts of “high” expert reviews, defined as above median scores (above 95, 17.5, and 95 for Parker, Robinson, and Vinous, respectively), as follows:

\[
\log A_{it} = \beta_1 H_{it} + \alpha_i + \delta_t + \epsilon_{it}
\]

\(A_{it}\) is the real quarterly average auction price for wine \(i\) in quarter \(t\). \(H_{it}\) is an indicator for if wine \(i\) had a high expert review affiliated with it at quarter \(t\).\(^{15}\) Lastly, \(\alpha_i\) are wine specific fixed effects and \(\delta_t\) are quarter fixed effects. We use this specification to investigate the reduced-form effects of Parker, Robinson, and Vinous reviews. We control for wine-specific linear trends as a robustness check.

<table>
<thead>
<tr>
<th></th>
<th>Log Mean Auction Price (Real)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Parker high score</td>
<td>0.144***</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
</tr>
<tr>
<td>Robinson high score</td>
<td>-0.028***</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
</tr>
<tr>
<td>Vinous high score</td>
<td>0.060***</td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
</tr>
<tr>
<td>Wine fixed effects</td>
<td>Yes</td>
</tr>
<tr>
<td>Quarter fixed effects</td>
<td>Yes</td>
</tr>
<tr>
<td>Wine-specific trends</td>
<td>No</td>
</tr>
<tr>
<td>(N)</td>
<td>10,109</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.959</td>
</tr>
<tr>
<td>Adjusted (R^2)</td>
<td>0.957</td>
</tr>
</tbody>
</table>

*Using robust SEs*

*** 1 % level; ** 5 % level; * 10 % level

Columns 1 and 4 illustrate results with Parker high score dummies, 2 and 5 Robinson high score dummies, and 3 and 6 Vinous high score dummies. At first, in columns 1-3, results seem to indicate that Robinson has a negative effect (-2.8%) on prices, while Parker and Vinous have strong positive effects (of 14.4% and 6%, respectively). However, once we control for wine-specific linear trends (a standard differences-in-differences robustness check when units receive treatment at different times) in columns 4-6, only Parker effects survive. Statistically significant at the 1% level, a high Parker score is affiliated with a 9.6% increase in auction price. In these results, Parker emerges as the key influencer, while Robinson and Vinous fall to the wayside.

These results demonstrate that scores (namely, Parker scores) are indeed associated with price movement (in ways that cannot otherwise be explained by wine or quarter fixed effects, or wine-specific linear time trends). However, reviewer scores are presumably correlated with quality, meaning we don’t know if the impacts are because the reviewer score movements are proxying the packets of quality information that are being released, or reflecting the pure reputational or publicity effects of the expert, or presumably both. In the next section, we discuss how to model...
the effects of expert reviews as a signal extraction problem. We then describe a timing restriction on the CellarTracker response which allows us to decompose the two via a VAR methodology. As such, we make use of a unique source of social media data to address the mechanism behind price response to expert ratings.\footnote{As such, we address the possibility, mentioned in Ashenfelter and Jones (2013), that expert ratings “influence prices because they create values that are independent of the function.”}

6 Signal Extraction Problem in the Case of High-End Wine

Expert and amateur reviews can be expected to arise in markets characterized by buyers who are asymmetrically and imperfectly informed concerning the quality of products. When these goods are relatively expensive (as with high-end wines) there is a strong incentive to seek and extract information from the noisy signals available concerning product quality. The public-good nature of information makes it natural for buyers to seek to share both their own evaluations of product quality as well as some version of the signals they have extracted from the noisy environment. These are the amateur reviews and are an additional source of information from which a signal could possibly be extracted.

6.1 Simple heuristic model

To develop our intuition and provide a basis for the identification strategy we employ, it is helpful to consider a simple model. We start with a simple non-dynamic single-market model, and then embed this within an adaptation of the finite-state signal-extraction model provided by Wallace (1992) who in turn was providing a simplified version of the classic signal-extraction model of Lucas (1972).

Let us begin by considering a market where the logarithm of supply price is a linear function of quantity $x$ to be auctioned, the quality of the wine $q$ and the reputation and other factors $\tau$:

$$p_s(x, \tau, q) = s_0 + s_1 x + s_2 \tau + s_3 q$$

This reflects the impacts on production costs of producing higher quality wines, as well as the costs associated with establishing and advertising a reputation or maintaining whatever other factors are represented by $\tau$.

The ‘true’ quality $q$ and reputation $\tau$ are important determinants of demand, and buyers are willing to bid higher prices for what they believe are higher quality wines from wine makers with established reputations. They have no direct observation of these, however, and must rely upon amateur reviews $\gamma$ (from CellarTracker or similar sources) and expert reviews $\varphi$ to provide noisy signals of quality and reputation.

The professional reviews depend on established reputations of the wine producer understood by the reviewer and taste-testing for quality:

$$\varphi(q, \tau) = p_0 + p_1 q + p_2 \tau + \epsilon_\rho$$

The amateur reviews are posted by wine enthusiasts who have sampled the product and are knowledgeable consumers. Their knowledge of wine maker reputation and other factors is less complete than the experts, but they are assumed to reflect these factors in their reviews by conditioning their signal on the log of prices from previous auctions. In the dynamic model this becomes an important consideration that allows us to solve for stationary equilibrium. For this heuristic model we have:

$$\gamma(p, q) = k_0 + k_1 p + k_2 q + \epsilon_\gamma$$

We assume that the error intrinsic to both the expert and the amateur reviews has mean zero so that $E[\epsilon_\rho] = E[\epsilon_\gamma] = 0$.

The log of price that buyers are willing to bid for wine depends on the quantity $x$ to be auctioned as well as whatever information about wine quality $q$ and wine maker reputation and other factors $\tau$ can be extracted from the noisy signals $\varphi$ and $\gamma$:

$$p_d(x, \varphi, \gamma) = d_0 - d_1 x + d_2 \varphi + d_3 \gamma$$
Equilibrium requires \( p_d(x, q, \gamma) = p_u(x, \tau, q) \). Solving and taking expectations we have an equilibrium price \( p^* \) with:
\[
\ln p^* = \beta_0 + \beta_1 \tau + \beta_2 q
\]
where:
\[
\beta_0 = \frac{d_1 s_0 + d_2 s_1 + d_3 s_2 + d_4 p_0 s_1}{d_1 + s_1 - d_2 k_1 s_1}
\]
\[
\beta_1 = \frac{d_2 p_0 s_1 + d_2 k_2 s_2}{d_1 + s_1 - d_2 k_1 s_1}
\]
\[
\beta_2 = \frac{d_3 k_2 s_1 + d_2 p_1 s_1 + d_1 s_2}{d_1 + s_1 - d_2 k_1 s_1}
\]
This simple framework, in which the log of equilibrium price is a linear function of two important but unobserved variables for which a noisy signal is available, is similar to the underlying structure of the model presented in Wallace (1992). That model involves agents who live for two periods and consume leisure \( \ell \) and a single consumption good that we (for obvious reasons) will refer to as wine or \( \omega \). Each agent maximizes:
\[
E[u_1(\ell) + u_2(\omega)]
\]
In the first period the agents are endowed with \( \lambda > 0 \) units of leisure. The economy has a constant returns to scale technology for converting leisure into wine, and the wine is only consumed by agents during the second period of their lives (after aging). The decision problem that must be made by the agents is to choose leisure consumption \( \ell < \lambda \) to consume when young, converting remaining leisure into wine \( \omega = \lambda - \ell \) that they consume when old.

The utility \( u_2(\omega) \) of the wine consumption \( \omega \) consumed at time \( t \) depends on the unobserved quality \( q_t \) and other factors \( \tau_t \) associated with or emerging from the wine production process, and these evolve over time. This dynamic evolution causes the agents to respond by changing the balance between leisure consumption \( \ell \) and wine production \( \omega \) and the price of wine \( p_t \) adjusts to clear the market. Agents maximize expected utility given in equation (9) by observing signals \( q_t \) that depends linearly on \( q_t \) and \( \tau_t \) plus an iid error term, and \( \gamma_t \) that depends linearly on lagged prices \( p_{t-1} \) and quality \( q_t \) plus an iid error term.

Coupled with assumptions that ensure that the utility sub-functions \( u_1 \) and \( u_2 \) are strictly concave and that quality \( q_t \) and other factors \( \tau_t \) are gross substitutes, we can apply an argument similar to Wallace (1992) to show that there is a unique solution to the signal extraction problem that produces a stationary outcome \( p_t \) that depends on \( q_t \) and \( \gamma_{t-1} \). Thus, in this context, a stationary price series is produced conditional on contemporaneous values of quality and lagged values of other factors or reputation, and this emerges from how agents extract the important information from the noisy signals, along with assumptions about the structure and stochastic properties of those signals.

7 Empirical Strategy

We use the intuition gained from the above model in conjunction with the recently developed panel structural vector autoregressive methods of Pedroni (2013) to decompose the signal contained in expert reviews into their component parts. We assume that the price of a high-end wine is influenced by both determination of the good’s quality and the remaining factors that impact supply and demand. While we model wine quality itself as evolving in a smooth manner, innovations to the information that arrives regarding its quality come in the form of shocks, as this information is made available at discrete moments in time. We call these innovations ‘quality information shocks.’ As the quality of wine evolves, wine drinkers are trying to figure out its quality. Shocks to the available information on a good arise as more people drink the wine and report on its quality. However, quality information shocks are not the only shocks to influence wine price. The signal extraction model posits that a stationary price series is produced conditional on contemporaneous values of quality and lagged values of other factors, where we think of “quality” as reflecting the consumer based notion of the quality of the liquid in the bottle and we think of the “other” factors as reflecting any other attributes associated with the bottle of wine that might impact the market price, including the reputation of any experts that have rendered opinions on the wine. These in turn are associated with noisy signals, for which both the amateur and expert reviews provide information. The purely stochastic components of these noisy signals are denoted by the orthogonalized vector of white noise shocks \( \epsilon_{it} = (\epsilon_{it}^p, \epsilon_{it}^q)' \).
As discussed in section 7.1 below, the identifying restriction that allows us to disentangle the quality signal shock, $\epsilon_{it}^q$, from the other shock, $\epsilon_{it}^p$, is that there is no direct immediate impact effect from $\epsilon_{it}^q$ to the average CellarTracker score in the period of the shock. This comes from the simple model dynamics due to the fact that $\gamma_t$ depends linearly on lagged prices $p_{t-1}$ and current quality $q_t$. Therefore, while price is a noisy signal that mixes liquid quality with other attributes that are valued, the combination of the crowd sourced amateur reviews with price allows us to disentangle the quality signal from the signal for the other attributes in a way that nevertheless allows amateurs to account for a temporally smoothed sense of the price of the wine into account.

Furthermore, in our signal extraction model, expert reviews have the potential to affect price in two ways: (1) they contain a quality information shock since reviewer expertise and knowledge is another innovation to the quality information stockpile, and (2) they disseminate information to a larger audience and larger base of consumers (due to the reviewer’s reputation), thus raising awareness of a given wine.\footnote{An expert reviewer’s dissemination effect differs from any potential CellarTracker dissemination effect since a much larger set of customers become aware of new Parker ratings than become aware of new CellarTracker reviews, given that CellarTracker reviewers are relatively anonymous wine enthusiasts. In this vein, we assume that the publicity effect from CellarTracker reviews is trivial.} A positive review from someone like Robert Parker is akin to a publicity event which raises total market demand by raising awareness of the product, but also, because it is in the form of an expert review, adds to the information stock. Thus, just as any shock could be decomposed into “quality information” and “other” shocks so too can expert review shocks. Note that this is also equivalent to decomposing our two signal shocks into components attributable to the expert versus those attributable to the rest of the market, such that we consider the further decompositions into $\epsilon_{it}^q = (\epsilon_{it}^{p^e}, \epsilon_{it}^{p^o})'$ and $\epsilon_{it}^p = (\epsilon_{it}^{e^e}, \epsilon_{it}^{e^o})'$. As such, we use explore the consequences of these decompositions by embedding the expert review scores as an event analysis within a structural identified panel VAR framework. In the next section we elaborate on the details of this approach.

### 7.1 Details of the Panel Structural VAR Approach

To see how the panel VAR framework will allow us to exploit the information from CellarTracker to identify the role of innovations to the amount of information available on the quality of wine and the role that this has on the price of wine we introduce here the details of the panel VAR set up. In particular, in keeping with the notation set out in Pedroni (2013), let $\Delta z_{it}$ be the vector of panel time series variables in their stationary form, observed over time periods $t = 1, ..., T_i$ for units $i = 1, ..., N_t$. Correspondingly, we take $\Delta z_{1, it}$ to be the CellarTracker review score averages for wine $i$ at time $t$, and similarly $\Delta z_{2, it}$ to be the natural log of the auction price average for wine $i$ at time $t$. The fact that the panel is unbalanced, with different start and end dates observed for different wines, and different numbers of wines observed for any given time period is reflected in the fact that the value $T_i$ is specific to wine $i$ and the value $N_t$ is specific to time period $t$. Since both the CellarTracker review scores and log auction price series evolve as stationary series, we enter their values in levels form for the $2 \times 1$ vector $\Delta z_{it}$. Furthermore, for ease of notation, we take $\Delta z_{it}$ to be the time demeaned variables so that fixed effects are automatically accommodated by this notation. Accordingly, for each wine, $i$, we can represent the potential dynamic relationship between the CellarTracker scores and log auction prices in a standard reduced form VAR as

$$\Delta z_{it} = R_{i1} \Delta z_{it-1} + ... + R_{iL_i} \Delta z_{it-L_i} + P_i + \mu_{it}, \quad (10)$$

or equivalently $R_i(L_i) \Delta z_{it} = \mu_{it}$. $R_i(L_i)$ is $I - \sum_{p=1}^{P_i} R_{i,j} L^j$, where the lag truncation $P_i$ is chosen by an information criteria to ensure that $\mu_{it}$ approximates a vector white noise process.

The challenge then is to be able to transform the estimates from these reduced form estimates into representations that are economically meaningful in terms of the quality versus other shocks discussed in section 6, namely $\epsilon_{it}$ where $\epsilon_{1, it}$ is the noisy signal shock for quality, $\epsilon_{it}^q$, and $\epsilon_{2, it}$ is the noisy signal shock for other non-quality shocks, $\epsilon_{it}^p$. Furthermore, this needs to be done in a way that is consistent not only with the heterogeneity in the dynamic responses as described in equation (10) above, but also takes into account the idea that the data is cross sectionally or “spatially” dependent across the various wines of the sample. This can occur for example when auction prices, or for that matter even CellarTracker reviews are responding to quality signals or other shocks that are common among the wines of the sample. Something as simple as changes to demand for high-end wines in response to macroeconomic conditions can generate such dependencies, as can
developments specifically in the wine market, such as the arrival of news regarding the quality of a future vintage, and so forth.

The strategy for addressing these challenges, as laid out in Pedroni (2013) is to exploit the structural VAR restrictions that are used to map from the reduced form shocks \( \mu_{it} \) into the orthogonalized economic shocks of interest, discussed in section 6.2, \( \epsilon_{it} = (\epsilon_{it}^{o}, \epsilon_{it}^{c})' \). This is accomplished by positing a factor model structure on the orthogonalized shocks such that

\[
\epsilon_{it} = \Lambda_{i} \hat{\epsilon}_{it} + \epsilon_{it}
\]

where the composite shock \( \epsilon_{it} \) is decomposed into a common shock \( \hat{\epsilon}_{it} \) and an idiosyncratic shock \( \epsilon_{it} \), with a diagonal loading matrix \( \Lambda_{i} \), which allows the relative importance of the common shocks to differ by wine unit. The fact that the loading matrix is diagonal implies that for example only common \( \epsilon_{it}^{c} \) shocks load into the composite \( \epsilon_{it}^{o} \) shocks and so forth, as one would expect for orthogonalized economic shocks. The fact that the factor structure is placed on the white noise orthonormal economic structural shocks rather than on the raw data implies that the loadings can be consistently estimated as the sample correlation between the common shocks \( \hat{\epsilon}_{it} \) and the composite shocks \( \epsilon_{it} \) for each wine \( i \), and can be estimated well with very short samples without the need for principle components estimation. Furthermore, Pedroni (2013) shows that the cross sectional averages of the panel data, \( \Delta \bar{z}_{it} = N_{t}^{-1} \sum_{i=1}^{N_{t}} \Delta z_{it} \) can be used to consistently recover estimates for the common shocks \( \hat{\epsilon}_{it} \).

The empirical approach then is to estimate the composite shocks via a structurally identified VAR applied to the data \( \Delta z_{it} \) for each wine individually and to estimate the common shocks via a structurally identified VAR applied to the cross sectional averages of the data \( \Delta \bar{z}_{it} \), which in turn allows estimation of the loadings, \( \Lambda_{i} \). In order to obtain the structurally identified VAR from the reduced form VAR representation of the data as per (10), one uses standard methods from the identified VAR literature. In particular, to construct impulse responses to our economic shocks, \( \epsilon_{it} = (\epsilon_{it}^{o}, \epsilon_{it}^{c})' \) we are interested to obtain the structural VMA representation of the data \( \Delta z_{it} = A_{i}(L) \epsilon_{it}, A_{i}(L) = \sum_{j=0}^{Q} A_{i,j} L^{j} \). Since our identifying restriction that allows us to disentangle the quality signal shock, \( \epsilon_{it}^{c} \); from the other shock, \( \epsilon_{it}^{o} \), is that there is no direct immediate impact effect from \( \epsilon_{it}^{o} \) to the average CellarTracker score in the period of the shock, this can be represented as a restriction on the short run impact matrix such that \( A_{i}(0)_{12} = 0 \forall i \), but which leaves further periods free to respond. The unique mapping from the reduced form VAR estimates \( R_{i}(L) \) and \( \mu_{it} \) to the structural form VMA components \( A_{i}(L) \) and \( \epsilon_{it} \) is then fairly standard, such that \( A_{i}(L) = R_{i}(L)^{-1} A_{i}(0) \) and \( \epsilon_{it} = A_{i}(0)^{-1} \mu_{it} \). In fact, since \( A_{i}(0) \) is upper triangular, this is the same mapping as a conventional Cholesky orthogonalization.

This mapping is computed individually once for each wine of the sample to obtain the composite shocks, and then once analogously for the cross sectionally averaged data to obtain the common shocks such that \( \hat{\epsilon}_{it} = \bar{A}(0)^{-1} \mu_{it} \) from the VAR estimated from the cross sectional averages, namely \( \bar{R}(L) \Delta \bar{z}_{it} = \mu_{it} \). The composite and common shocks are then used to estimate the loadings \( \Lambda_{i} \), which are then used to construct the wine specific responses to the common and idiosyncratic versions of the shocks\(^{18}\), namely

\[
\bar{A}_{i}(L) = A_{i}(L) \Lambda_{i}, \quad \bar{A}(L)_{i} = A_{i}(L)(I - \Lambda_{i} \Lambda_{i}')
\]

It is the distribution of the wine specific responses \( \bar{A}(L)_{i} \) to the idiosyncratic structural shocks that is studied in the subsequent section, as these are the responses to the orthogonalized economic shocks of interest that have been controlled for cross sectional dependencies across the panel of wine prices and CellarTracker scores.

A further novelty of the approach we take here is that we embed within the heterogeneous panel VAR estimation and inference what is effectively a panel event analysis, wherein changes to the various expert reviews are treated as events with known timing, but unknown consequences. This is accomplished in a manner analogous to event analysis in time series analysis, whereby the event is entered into the VAR estimation as a dummy, \( d_{it} \), which takes the value 0 up to the point of the event, and then takes on the value of event from the point in time at which it occurs onward through the remainder of the sample. In particular, the value of the event is taken to be the change in the expert score, so that for example if the expert score prior to the beginning of the sample was 91, but was subsequently changed to 93, the event would take the value 0 up to the point of the score change, at which point it would take the value 2, until any further score change occurred.

\(^{18}\)See Pedroni (2013) for further details
Analogous to the way the individual wine data is treated, the cross sectional average of the dummy events $d_t = N_t^{-1} \sum_{i=1}^{N_t} d_{it}$ is included in the cross sectional VAR estimation and the raw event dummies $d_{it}$ are included in the composite VAR estimation for each of the individual wines. The mapping from the reduced forms to the structural forms then allow us analogously to decompose the event effects into components that are analogous to quality signal shock $\epsilon_{it}^\rho$ and the other shock, $\epsilon_{it}^\kappa$, so that the dynamic impact of the expert reviews can be composed into these two structural components. It is this decomposition that we also further study in the next section.

8 Results

We present results from the panel structural vector autoregressive method based on our previously described identification strategy. The graphs display impulse responses, which describe the reaction of the endogenous variables (CellarTracker scores and log auction prices) at the time of shocks (quality information and other) and during subsequent time periods (quarters). There exist impulse responses for each wine, each of which is different. The figures represent responses for the sample of wines by displaying the interquartile range (that is, the range between the 25th and 75th percentile responses), also marking the median (black bar) and mean (red point). We first present results on the baseline VAR and endogenous variable responses to quality information and other shocks based on our previously outlined Cholesky decomposition. Second, we then present how our endogenous variables respond to expert reviews (Parker, Robinson, and Vinous reviews). Finally, we get into the crux of the signal extraction problem and structurally decompose those responses to expert reviews into components associated with quality information and other shocks.

8.1 Baseline VAR

Figure 4: Baseline VAR Responses

We present impulse responses of CellarTracker scores and auction prices to quality information and other shocks in Figure 4. We always report responses to idiosyncratic shocks to control for cross-sectional dependencies.

---

19 Relevant VAR code was run in RATS.
20 Confidence intervals will be added to results in the future. They require a bootstrap, which requires more time.
21 For wines to be included in each expert VAR analysis (steps 2-3), they must have an expert review which occurs sometime after the first 8 periods but before the last 4 periods. This leaves 115 wines with such Parker reviews, 136 for Robinson, and 61 for Vinous. However, the baseline VAR (step 1) is meant to be generated using the full sample of 355 wines. For this paper draft, it is generated using the subsample of Parker wines.
22 We always report responses to idiosyncratic shocks to control for cross-sectional dependencies.
23 This figure uses the subsample of 155 wines with adequately timed Robert Parker reviews. In the future, we
quarter response of CellarTracker scores to other shocks to be 0. We rescaled quality and hype shocks so that they can be reinterpreted in terms of the units of the response variables. Specifically, we scale the quality information shock such that it is reinterpreted as a quality information shock that increases CT scores by 1.0 in the quarter in which the shock occurs, and we rescale the other shock such that it is reinterpreted as an other shock that increases auction prices by 1% in the quarter in which it occurs. The rescaling has not changed the estimation; it simply changes the interpretation of counter-factual that we are investigating based on the estimates.

### 8.2 Expert Shocks

Now that we have presented responses to shocks in the baseline VAR, we compare responses specific to expert reviewers. As mentioned in section 7, recall that the value of the event is taken to be the change in the expert score. Figure 5 illustrates the response of CellarTracker scores and log auction prices to a one-unit change for each of the three reviewers.

![Figure 5: Responses to Expert Shocks](image)

Looking at the means for the first period response, Parker moves auction price more than his competition – a one unit increase in his scores means an average 0.2% increase in auction prices. However, in the second period the mean response is negative and then goes to zero. Meanwhile, responses to Robinson stay positive in the mean over the 6 quarters, while for Vinous they become and stay negative starting in the quarter after the review. The magnitudes of the reviewer effects in all three above are understandably smaller here than the differences-in-differences results because here we use a unit change reviews, rather than an indicator for “high” versus “low” score.

Meanwhile, CellarTracker responses display much heterogeneity across wines in their responses to expert reviews. Robinson displays the largest interquartile range, followed by Vinous, and then Parker. Considering that Robinson is perhaps less of an American influencer than the other two, this is understandable. The graph shows that for many wines social media reviews often decrease in response to favorable expert reviews. In these cases, there could be a negative effect from the publicity generated by an expert review that could outweigh the positive effect of the favorable potential underlying quality signal. This evidence could be suggestive of dynamics behind

---

24 Since Robinson scores out of 20, the changes are often bigger for her than for Parker or Vinous.
ratings conditional on reference points, defined as expectations in part defined by well-know expert reviewers.

8.3 Expert Shocks: Decomposed

By using the Cholesky decomposition previously discussed, we can orthogonalize the quality information and other shocks in the context of the review events. These results are presented in Figures 6 and 7. In a sense these figures are the product of Figures 4 and 5, as the previous graphs show the responses to expert shocks before decomposition and the responses to generic quality information and other shocks. Again, the identifying restriction limits the first quarter response of CellarTracker scores to review other shocks (in this context, publicity shocks) to be 0, as seen in the upper-right graph.

Figure 6: CellarTracker Responses to Decomposed Expert Shocks
CellarTracker Impulse Responses to Quality Information Expert Shocks

Figure 7 shows a very different picture by reviewer. For Parker, the auction price response to other expert shocks features an interquartile range all above zero. The mean response is larger than for the other experts. Meanwhile price responses to Parker other shocks have a mean of 0 and are even meaningfully negative in the following quarter. These results imply that Robert Parker’s reviews influence the market primarily due to the publicity effect that accompanies such reviews rather than due to their quality signal.

However, responses to Robinson quality information shocks stay positive in the mean throughout the 6 quarters, while Vinous quality information shocks are very high in the original quarter and then quickly become negative in the mean. Distributions of wine price responses skew much more negative in response to other expert shocks for Robinson and Vinous. Therefore, results suggest that the power of the publicity channel is unique to Robert Parker’s reviews.

Given Robinson and Vinous’s lack of influence in our original reduced-form estimates, it is not clear how meaningful it is to speak to the mechanism behind their influences. Meanwhile, given Parker’s clear influence shown in section 5, it is meaningful to show that his effect is mainly through the publicity channel. In the end, our results point to differential impacts on prices by reviewer as well as differential importance of quality information and publicity by reviewer.
8.4 Next steps

A future part of our analysis will be to ask what characteristics of the wine types cause the market to “play with” the expert versus “play against” the expert on either the quality information or other components. The previous section graphically highlights the heterogeneity of responses across different wines. We can investigate this heterogeneity across geographic lines as well as across wine characteristics by using collected text that accompanies CellarTracker and expert reviews. In effect, wines characteristics can be identified through text analysis and used to explore the context for the displayed heterogeneity. This opens up a new dimension of data to investigate in future drafts.

9 Conclusion

In studying experts in the market high-end goods, we consider the case of wine. We first generate reduced-form estimates of expert reviews on prices and find evidence that Robert Parker is unique in his influence on prices when compared alongside his taste-making peers Jancis Robinson and Antonio Galloni. We then address the mechanism underlying how experts may influence prices through development of a signal extraction model and use of structural vector autoregression (VAR) (paired with a timing restriction on amateur social media reviews) to separate out the quality information from the publicity components of expert reviews. The SVAR approach provides us with suggestive evidence that Robert Parker influences prices primarily through the publicity channel, rather than the quality information channel. These results are consistent with Gibbs, Tapia and Warzynski (2009), which discussed the possibility that an increase in the proportion of naive consumers was responsible for the increasing influence of expert wine reviews on price. The results also respond inherently to Ashenfelter and Jones (2013)’s finding that experts have some affect on price independent of quality. Our results suggest that expert reviews in high-end goods markets can indeed create values independent of the a good’s function.

Our paper suggests that we have not reached the end for “the expert” in the realm of high-end goods when it comes to the market’s determination of price. However, with the growing popularity of crowd-driven technologies such as CellarTracker – to borrow from the gist of Rosen (1981)’s

---

25 We reference a potential shift in the evaluation of goods discussed by Arora and Vermeylen (2013)
discussion of the influence of economic progress on value— who knows “what changes will be wrought by cable, video cassettes and home computers?” More aptly updated for the modern day and for our specific context, who knows what changes will be wrought by social media’s expanding presence? It is possible that, as Arora and Vermeylen (2013) puts it, “crowd wisdom” will become “the new guide in constructing and evaluating knowledge,” as well as measures and signals of quality, even in the realm of high-end goods that previously strictly required the advice and opinions of renown experts. Regardless, it is no question that the roles of social media and experts will meaningfully evolve in the coming years, though their trajectories might be harder to predict than that of a bottle of Bordeaux.

26 The effects in this model seem applicable to wine; he “similarly argues that some goods have a public good aspect of joint consumption.”

27It is painfully obvious what three decades of technological advances have “wrought” these examples to be woefully outdated.
References


McAuley, Julian John, and Jure Leskovec. 2013. “From amateurs to connoisseurs: modeling the evolution of user expertise through online reviews.” 897–908, ACM.


