Let the Market Be Your Guide: Estimating Equilibria in Differentiated Product Markets with Class-Membership Uncertainty

Marco Costanigro  
*Dept. of Agricultural and Resource Economics*  
*Colorado State University*

Jill J. McCluskey and Ron C. Mittelhammer  
*School of Economic Sciences*  
*Washington State University*
Wine Differentiation

- Wines are characterized and valued along widely varying attribute dimensions

- “Similar wines” are near substitutes and competitors in the consumption decision

- Beyond some level of differentiation, wines may no longer be viewed as part of the same product class, nor viewed as near substitutes or competitors

- Attributes can be valued differently across different product classes
  (e.g., the effect of aging on $2 Chuck vs. Penfolds Grange)

Which wines belong to what product classes and submarkets?
How do attribute evaluations differ between product classes?
Previous Wine Studies

1. Combris et al. (1997, 2000) showed that when regressing objective and sensory characteristics on wine price, the objective cues (such as expert rating score and vintage) are significant, while sensory variables (such as tannins content and other measurable chemicals) are not.

2. Oczkowski 1994; Landon and Smith 1997; Schamel et al. 2003, Angulo et al. 2000 indicate that ratings by specialized magazines are significant and should be included when modeling wine prices.

3. Oczkowski (2001) finds that tasting scores are proxies for quality

4. Angulo et al., 2000; Schamel and Anderson, 2003 find that region of production, the collective reputation of the district, and the vintage are significant variables.
Previous Wine Studies

5. Thrane (2004) indicates it is unlikely that the same hedonic function will apply for red and white wines.

6. Both the wine industry (Ernst and Young, 1999) and typical wine consumers (Hall et al., 2001) use price categories to define product-class categories.

7. Ernst and Young (1999), based on qualitative survey information, divided wines into commercial, semi-premium, premium and ultra-premium categories on the basis of retail price ranges.

8. Hall et al. (2001) found that price is used as a quality cue and that consumers look for different attributes, or value the same attributes differently, depending on the occasion the wine is meant to be consumed.
**Previous Wine Studies**

**Past Econometric Research:** relationship between wine prices and attributes

- Substantial agreement on some determinants (attributes) of price in the hedonic function:

  \[ P = f(\text{aging, quality, variety, region of production, vintage}) \]

- Economists generally estimated a single hedonic function for wine

**Research questions**

- Should analysts estimate multiple hedonic functions for wine?
- Can we determine 4 product classes/submarkets for wine?
Previous Evidence: Price Segmentation

- **Costanigro et. al. (Journal of Agricultural Economics, forthcoming)**
  - Estimated 4 price-range specific hedonic functions
  - Product classes: Commercial, Semi-Premium, Premium and Ultra-Premium
  - Determined price range partitions by minimizing the SSE of the overall model

- **Results**
  - Implicit prices of attributes differ significantly across price ranges
  - Substantially increased explanatory power of the segmented model vs. pooled model

- **Limitations**
  - Price might not be the only variable defining wine classes
  - Wine classes might be overlapping in price
Methods

• **Objective:**
  - Estimate class specific hedonic functions:

  \[ y_i = g_j(x_i; \beta_j) + \varepsilon_i \]

  when class membership \( j \in \{1, 2, ..., J\} \) is *uncertain* for observations \( i = 1, ..., n \)

• **Desirable properties of any methodology used**
  - Classes determined within the model estimation procedure
  - Use multiple determinants of class membership
  - Allows for price overlap

• **Existential approaches:**
  1. Likelihood methods: latent class or finite mixture models, random parameters
  2. Clustering approaches
Local Polynomial Regression Clustering (LPRC)

**Intuition**

1. Estimate the hedonic function locally and nonparametrically via **local polynomial regression**
   - Results in *n observation-specific* estimates of the hedonic function

2. Group observations into *J* groups using a **clustering algorithm** based on finding similarities among the *n* hedonic functions in terms of the effects of attributes on implicit prices

3. Estimate *J* simple, class-specific parametric hedonic functions via least squares.
**LP**

**RC**

**Ratiocination**

\[ y_i = \sum_{j=1}^{J} g_j(x_i; \beta_j)I_{(D_j)}(x_i, z_i) + \varepsilon_i \]

- where \( \bigcup_{j=1}^{J} D_j = D \) is a partition of the sample data
- Identify partition of the observed data such that the estimated model

\[ y_i = \sum_{j=1}^{J} \hat{g}_j(x_i; \hat{\beta}_j)I_{(\hat{D}_j)}(x_i, z_i) + \nu_i \]

approximates well the relationship between \( y_i \) and \( x_i \) up to the \( r \)th order derivative relationship.
LPRC

*Ratiocination*

- Use local polynomial regression of order \( r \) to generate local nonparametric observation-specific estimates of the derivatives
  
  \[
  \hat{\mathbf{b}}^r (x_0) = \begin{bmatrix}
  \frac{\partial \hat{g}(x)}{\partial x} \bigg|_{x_0} & \frac{\partial^2 \hat{g}(x)}{\partial x^2} \bigg|_{x_0} & \ldots & \frac{\partial^r \hat{g}(x)}{\partial x^r} \bigg|_{x_0}
  \end{bmatrix}
  \]

  for each sample point \( x_0 \).

- Cluster the derivative values into \( J \) classes on the basis of similarity in the values of \( \hat{b}^r (x_0) \).

- The clustering step identifies data partitions for which the functional relationship between \( y \) and \( x \) is relatively stable to the \( r^{th} \) order.
**Empirical Implementation**

**Step 1: Local Polynomial Regression (1\textsuperscript{st} order in this application)**

\[
\hat{b}(x_0; x) \equiv \underset{a,b}{\text{arg min}} \left\{ \sum_{i=1}^{n} [y_i - a - (x_i - x_0)b]^2 K((\xi_i - \xi_0')/h) \right\}
\]

- **Kernel Weighting Function (Tri-Cube)**

\[
K((\xi_i - \xi_0')/h) = W[((\xi_i - \xi_0)(\xi_i - \xi_0'))^{1/2}/h] = W_{i0}
\]

where \( W[\mu] = \begin{cases} (1-|u|^3)^3 & \text{for } |u| \in [0,1] \\ 0 & \text{elsewhere} \end{cases} \)

- **Store observation-specific first derivatives:**

\[
\left\{ \hat{b}(x_i; x), \; i = 1, \ldots, n \right\}
\]
• **Step 2: Ward Clustering**
  
  - Group observations on the basis of similar $\hat{b}(x_i; x)'s$
  
  - Ward algorithm minimizes the Deviation SS
    
    $$DSS_j = \sum_{i=1}^{N} (\hat{b}(x_i; x) - \hat{b}_j)'(\hat{b}(x_i; x) - \hat{b}_j)$$
    
    for a given number of clusters, $J$

• **Step 3: Cluster specific regression (OLS)**
  
  - Obtain class specific hedonic models
    
    $$y_i = x_i \beta_j + \varepsilon_i, \; i \in D_j, \; j = 1, \ldots, J$$
Data: Overview

- **10 years of observations** (1991-2000) for CA and WA **red wines** from *Wine Spectator* Magazine.

- **9,820 observations** (8,848 from California)

- **Variables include:**
  - Price
  - *Wine Spectator* score
  - Years of aging before commercialization
  - Number of cases produced
  - Seven CA macro-regions of production (*Napa Valley, Bay Area, Sonoma, South Coast, Carneros, Sierra-Foothills, and Mendocino*) plus **CA Generic** and **Washington** (Wine Spectator Designations)
  - Variety (Zinfandel, Pinot Noir, Cabernet, Merlot, Syrah, blends)
  - Vintage
  - Patronage: “reserve”, “estate produced” and/or “vineyard name”
Data: American Viticultural Areas (A.V.A.) in California and Washington State

- Non A.V.A: Generic California, Washington
- Macro A.V.A: Bay Area, South Coast, Sierra
- Micro A.V.A: Napa Valley, Sonoma, Mendocino, Carneros
Results

1. Classification results
   - Partition the data into product classes
   - Identify characteristics of each wine class

2. Hedonic valuation results
   - How wine attributes are valued
   - How attribute valuation changes across classes
Classification Results

Score Distribution by Class

Tasting Score
Classification Results:
Region of Production

Commercial
- Mendocino, 6.3%
- Generic Ca, 59.6%
- Sierra, 7.6%
- South Coast, 3.4%
- Bay, 2.9%
- Sonoma, 9.2%
- Napa, 2.8%
- Wa, 7.6%

Semi Premium
- Mendocino, 6.5%
- Generic Ca, 9.5%
- Sierra, 4.2%
- Carneros, 5.3%
- South Coast, 11.3%
- Bay, 5.9%
- Napa, 14.0%
- Wa, 13.9%
- Sonoma, 28.6%

Premium
- Mendocino, 2.6%
- Generic Ca, 1.0%
- Sierra, 0.4%
- Carneros, 5.5%
- South Coast, 9.9%
- Bay, 6.6%
- Wa, 9.6%
- Napa, 36.7%
- Sonoma, 27.8%

Ultra Premium
- Mendocino, 2.6%
- Generic Ca, 1.1%
- Sierra, 1.6%
- Carneros, 4.2%
- South Coast, 9.4%
- Bay, 4.6%
- Sonoma, 21.0%
- Wa, 5.9%
- Napa, 49.6%
Classification Results: Grape Variety

- **Commercial**
  - Zinfandel: 30%
  - Syrah: 4%
  - Merlot: 23%
  - Pinot: 9%
  - Cabernet: 32%
  - Blend: 2%

- **Semi Premium**
  - Zinfandel: 32%
  - Syrah: 6%
  - Merlot: 22%
  - Pinot: 12%
  - Cabernet: 20%

- **Premium**
  - Zinfandel: 11%
  - Syrah: 9%
  - Merlot: 18%
  - Cabernet: 30%
  - Blend: 7%

- **Ultra Premium**
  - Zinfandel: 10.6%
  - Syrah: 0.6%
  - Merlot: 11.2%
  - Pinot: 17.0%
  - Cabernet: 43.0%
  - Blend: 13.2%
Results: Implicit Prices

|$/Point$

Tasting Score

Commercial  Semi Premium  Premium  Ultra Premium

95% C.I.

Same Attribute

Commercial
Semi Premium
Premium
Ultra Premium

95% C.I.

Same Attribute
Results: Implicit Prices

Premium/Discount

Region of Production

Benchmark: Generic California
So What? Conclusions

1. A new broadly applicable methodology (LPRC) to estimate class specific parametric models under class uncertainty was developed.

2. Four wine classes were clearly identified and characterized.
   - The categorization can be used to identify competitors in a given wine market (within the US or in the world wine market).

3. Implicit pricing of the attributes was shown to be different across classes.
   - Information on implicit prices can be used by wine producers for evaluating attribute mix by product class.

4. Wine marketing, consumer and demand analysis should be specific to wine class.
Future Research

1. LPRC:
   - Make number of classes endogenous
     - Are there more than 4 product classes for wine?
   - Reduce computational cost of estimation procedure
   - Monte Carlo simulations to investigate finite sample categorization and estimation performance

2. Wine economics
   - Are AVA price premia changing over time?
   - Macro AVA vs Micro AVA. Which matters most?
   - Is the increasing number of AVAs lowering their ability to act as signal of quality?
Finis
(for Now)

Two Buck Chuck
Appendix II More Results From LPRC
Classification Results

Aging Distribution by Class

Aging Years

- Commercial: 2.31 (4), 2.65 (5), 2.86 (8)
- Semi-Premium: 1.00 (1), 1.10 (3), 1.20 (2)
- Premium: 1.00 (1), 1.10 (3), 1.20 (2)
- Ultra-Premium: 1.00 (1), 1.10 (3), 1.20 (2)
Classification Results

Cases Produced by Classes

Cases Produced

Logarithmic Scale
### Results: Implicit Prices

<table>
<thead>
<tr>
<th>Attribute</th>
<th>$/100 Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commercial</td>
<td></td>
</tr>
<tr>
<td>Semi Premium</td>
<td></td>
</tr>
<tr>
<td>Premium</td>
<td></td>
</tr>
<tr>
<td>Premium</td>
<td></td>
</tr>
<tr>
<td>Ultra Premium</td>
<td></td>
</tr>
</tbody>
</table>

#### Notice:
1) Estimates are different
2) Significantly so (non-overlapping C.I.)
Results: Implicit Prices

Grape Variety

Benchmark: Zinfandel