Who wants some bubbly?

Drivers of Champagne Consumption in the US

Applied Regression Final Project

Rohit Dhingra, David J Wolkoff,
Stephanie Harvell, Neeta Boddapati,
and Liz Harrison

December 2011

Who wants some bubbly?

Context

- Connected with CBS EMBA alum Michelle who works at boutique (~1% market share) champagne company Laurent-Perrier
  - Champagne industry immature in terms of sales forecasting (especially compared to other beverage categories)
  - LP unable to predict sales, resulting in inventory issues

Project Description & Goals

- Conduct a regression analysis to determine the key drivers (lagging variables) of champagne consumption in the US
- Enable Michelle to accurately predict champagne sales for her company based on observable, historical variables

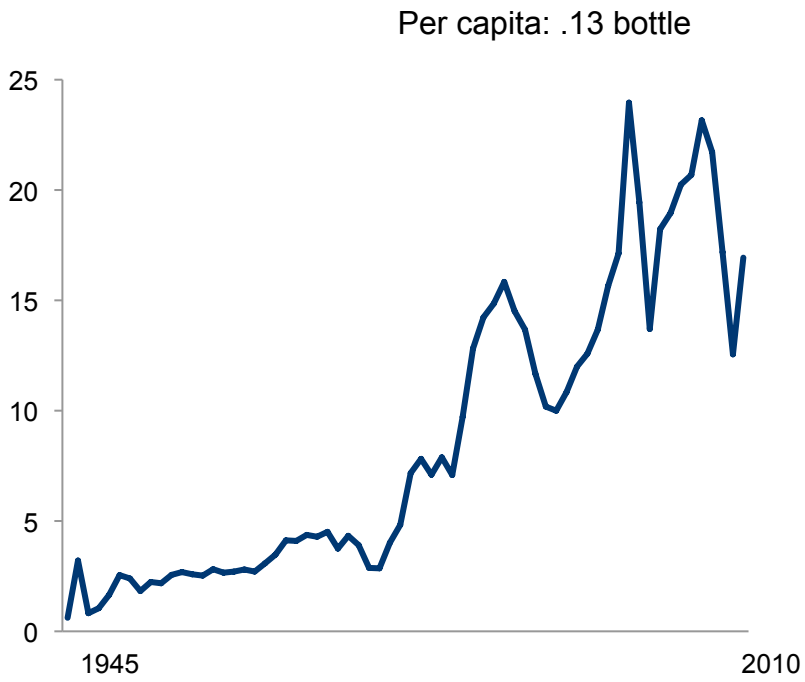
DEPUIS 1812 SINCE
Laurent-Perrier
CHAMPAGNE

Champagne Industry Overview

- Very small proportion of US alcohol market by value (1.9%) and volume (0.5%)
- Vast majority consumed during holiday season
- One of the biggest losers during the recession declining by 24% between 2006 and 2009 (market value) as consumers switched to more affordable alcohols
- Dominated by LVMH Moet Hennessy Louis Vuitton (63% market share by volume)

Alcohol Type	Volume %	Value %
Beer	81.8	49.3
Spirits	6.0	28.3
Wine	9.6	18.2
RTDs	1.9	2.1
Champagne	0.5	1.9

Champagne Consumption
Million 750ml Bottles



Source: Euromonitor; Client Data

Methodology

1. Project Kick-Off & Variables Considered

- Conducted kick off phone call to discuss what client wants to learn: What are the key drivers of champagne sales?
- Brainstormed potential X variables with client & received information regarding pricing, shipment of cases to the US, etc (limited data available)
- Hypothesized about other possible trend based predictors: demographic trends, alcohol sales, holiday attributes, etc.

2. Variable Correlation – Controlling for Multicollinearity

- Observed correlation between independent variables using correlation matrix

3. Identifying the Key Variables – Minitab Best Subsets Regression

- Conducted mini tab best subsets analysis to predict which variables were most meaningful in predicting champagne sales
- Removed variables that were not found to be significant

4. Full Regression Output – What does it all mean?

- Conducted full regression in excel to create the best fit model to help predict champagne sales

Appendix: Exponential Smoothing

- Created an exponential smoothing model with both “level” and “trend” terms
- Compared the sum of square errors to regression and determined regression was better

1. Variables Considered

Dependent: Champagne consumption per capita (750ml bottle)

Independent:

- Total Sales (volume x price)
- Price per bottle in Euros
  - Current
  - Constant (inflation adjusted)
- US GDP
- Dow Jones Industrial Average
- US Unemployment Rate
- Average Household Income in 2010 USD
  - Top 5%
  - Quartiles
- US Population
  - Total
  - By Age (focus on >21)
- Alcohol Consumption
  - Distilled Spirits (gallons and per capita)
  - Beer (gallons and per capita)
  - Wine (gallons per capita)
- # of Days between Thanksgiving and New Year's Day
- # of European tourists visiting the US

*** Note: All variables were lagged by one period*

2. Variable Correlation

- Using Excel's correlation table function we created a correlation matrix of all the independent and dependent variables
- In order to help avoid multicollinearity we removed independent variables which were highly correlated with other independent variables
- When a set of variables was highly correlated with one another we decided upon which variable to keep based on how well it correlated with the dependent variable and how poorly it correlated with other independent variables
- After the process we were left with 10 independent variables

Excel Correlation Matrix

Correlation Matrix	Per capita consumption	Euros Bottle (constant)	Euro Tourists to the US ('000)	US GDP	US Unemployment	Top 5% Avg Household Income (2010 USD)	Top 25% Avg Household Income (2010 USD)	Second 25% Avg Household Income (2010 USD)	Third 25% Avg Household Income (2010 USD)	Bottom 25% Avg Household Income (2010 USD)	DJIA	US Population	Distilled spirits (per capita)	Beer (per capita)	Population, 20-29	Population, 30-39	Population, 40-49	Population, 50-59	Population, 60+	Total Wine per Resident (in gallons)	Time
Per capita consumption	1.00																				
Euros Bottle (constant)	0.43	1.00																			
Euro Tourists to the US ('000)	0.55	0.25	1.00																		
US GDP	0.62	0.18	0.79	1.00																	
US Unemployment	-0.48	-0.42	-0.07	0.05	1.00																
Top 5% Avg Household Income (2010 USD)	0.82	0.50	0.75	0.86	-0.32	1.00															
Top 25% Avg Household Income (2010 USD)	0.85	0.52	0.73	0.83	-0.38	0.99	1.00														
Second 25% Avg Household Income (2010 USD)	0.85	0.61	0.65	0.68	-0.55	0.95	0.97	1.00													
Third 25% Avg Household Income (2010 USD)	0.83	0.63	0.49	0.50	-0.71	0.83	0.87	0.96	1.00												
Bottom 25% Avg Household Income (2010 USD)	0.78	0.69	0.48	0.45	-0.72	0.82	0.85	0.94	0.98	1.00											
DJIA	0.78	0.51	0.80	0.88	-0.33	0.95	0.94	0.89	0.77	0.74	1.00										
US Population	0.61	0.23	0.77	0.99	0.06	0.88	0.85	0.71	0.52	0.48	0.89	1.00									
Distilled spirits (per capita)	-0.07	-0.28	-0.12	0.25	0.30	-0.09	-0.09	-0.20	-0.18	-0.27	-0.01	0.19	1.00								
Beer (per capita)	-0.51	-0.25	-0.80	-0.74	-0.11	-0.75	-0.71	-0.61	-0.39	-0.41	-0.71	-0.77	0.38	1.00							
Population, 20-29	-0.65	-0.38	-0.78	-0.74	0.20	-0.87	-0.84	-0.79	-0.62	-0.64	-0.83	-0.78	0.46	0.91	1.00						
Population, 30-39	-0.60	-0.18	-0.70	-0.98	-0.07	-0.83	-0.81	-0.67	-0.51	-0.45	-0.85	-0.98	-0.37	0.64	0.64	1.00					
Population, 40-49	0.68	0.47	0.55	0.52	-0.46	0.79	0.79	0.81	0.70	0.74	0.74	0.58	-0.59	-0.73	-0.92	-0.44	1.00				
Population, 50-59	0.66	0.25	0.75	0.99	0.00	0.90	0.88	0.75	0.58	0.54	0.90	0.99	0.22	-0.74	-0.76	-0.99	0.58	1.00			
Population, 60+	-0.18	-0.33	0.34	0.47	0.66	0.03	-0.01	-0.19	-0.31	-0.38	0.09	0.41	0.68	-0.15	0.11	-0.50	-0.47	0.38	1.00		
Total Wine per Resident (in gallons)	0.53	0.04	0.57	0.89	0.11	0.68	0.67	0.54	0.42	0.33	0.71	0.86	0.61	-0.43	-0.39	-0.94	0.18	0.89	0.63	1.00	
Time	0.61	0.20	0.80	0.99	0.07	0.88	0.85	0.70	0.51	0.47	0.88	1.00	0.16	-0.80	-0.79	-0.97	0.58	0.99	0.43	0.85	1.00

3. Identifying Key Variables

Minitab Output

Vars	R-Sq	R-Sq(adj)	Mallows CP	S	Price/Bottle	Euro Tourists to US	US GDP	Unemployment	2nd 25% Household	Dow Jones	Spirit/capita	Beer/capita	Wine/capita	Time
1	73.0	71.6	5.7	0.017359					X					
1	61.3	59.4	15.9	0.020765						X				
2	74.5	71.8	6.4	0.017313	X				X					
2	73.9	71.1	6.9	0.017512					X		X			
3	76.4	72.5	6.7	0.017095			X		X					X
3	75.0	70.8	8.0	0.017613	X				X		X			
4	79.3	74.4	6.2	0.016480			X		X			X		X
4	77.6	72.3	7.7	0.017150			X		X		X			X
5	81.6	75.9	6.1	0.016007			X	X	X		X			X
5	80.5	74.4	7.1	0.016492		X	X		X			X		X
6	84.8	78.7	5.3	0.015034	X		X	X	X		X			X
6	84.4	78.2	5.7	0.015233	X		X	X	X				X	X
7	86.3	79.4	6.1	0.014801	X	X	X	X	X		X			X
7	85.7	78.5	6.6	0.015119	X		X	X	X	X			X	X
8	87.0	79.1	7.4	0.014919	X	X	X	X	X	X	X			X
8	86.7	78.5	7.7	0.015116	X	X	X	X	X	X			X	X
9	87.4	77.9	9.1	0.015318	X	X	X	X	X	X	X		X	X
9	87.3	77.8	9.2	0.015374	X	X	X	X	X	X		X	X	X
10	87.5	76.1	11.0	0.015940	X	X	X	X	X	X	X	X	X	X

- Using minitab's best subset tool, we iteratively removed variables until we were able to run an error-free simulation
- Selected model with the highest R-Sq, and lowest Mallows CP and S
- Best model includes 7 key variables, removing:
  - Dow Jones
  - Beer consumption per capita
  - Wine consumption per capita

4. Full Regression

- We performed a regression of the 22 observations with sales as the dependent variable and using the 7 parameters identified in the previous steps as the independent variables
- We were able to obtain a regression equation with an adjusted R-square value of 0.79
- Key findings include:
  - Price has an inverse effect on sales showing negative price elasticity as expected
  - We believe that distilled spirits consumption has a negative coefficient as spirits can cannibalize wine sales
  - The time variable suggest that bottles sold per capita is trending downwards

Regression equation and output

Regression Equation:

Per capita sales =
-1.252 - 0.005 x Euros per bottle
- 4.631 x 10^{-6} x Euro Tourists to the US (‘000)
+ 1.036 x 10^{-13} x US GDP + 2.259 x US Unemployment
+ 2.187 x 10^{-5} x Second 25% avg. household income in 2010
- 0.304 x Distilled spirits (per capita)
- 0.051 x Time

	Coeff.	*Std Errors*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-1.252	0.321	-3.902	0.002	-1.940	-0.564	-1.940	-0.564
Euros Bottle (constant)	-0.005	0.003	-1.844	0.086	-0.011	0.001	-0.011	0.001
Euro Tourists to the US ('000)	0.000	0.000	-1.216	0.244	0.000	0.000	0.000	0.000
US GDP	0.000	0.000	3.313	0.005	0.000	0.000	0.000	0.000
US Unemployment	2.260	0.820	2.756	0.015	0.501	4.018	0.501	4.018
Second 25% Avg Household Income (2010 USD)	0.000	0.000	4.449	0.001	0.000	0.000	0.000	0.000
Distilled spirits (per capita)	-0.304	0.111	-2.738	0.016	-0.542	-0.066	-0.542	-0.066
Time	-0.051	0.015	-3.350	0.005	-0.083	-0.018	-0.083	-0.018

Appendix: Exponential Smoothing

- To see if we could find a better way to predict per capita sales we tried to model the data with exponential smoothing
- Due to the large trends seen in the data we included a “trend” term in addition to a “level” term
- Using excel’s solver we found the optimal “alpha” and “beta” coefficients for our “level” and “trend” terms
- After finding that the sum of the square differences between the actual data and the forecast were greater than that of our regression we decided to stick with the regression model instead

Exponential Smoothing Results

Alpha
0.2927

Beta
0.6407

Sum Sq Errors
0.0110

Year	Per Capita Sales	Level	Trend	Forecast	Error^2
1989	0.106	0.106	0.000		
1990	0.091	0.102	-0.010	0.106	0.000246
1991	0.079	0.095	-0.011	0.092	0.000159
1992	0.078	0.090	-0.005	0.084	0.000042
1993	0.084	0.088	0.002	0.085	0.000001
1994	0.093	0.090	0.007	0.091	0.000006
1995	0.098	0.092	0.005	0.096	0.000002
1996	0.106	0.096	0.007	0.098	0.000078
1997	0.122	0.104	0.013	0.104	0.000327
1998	0.133	0.112	0.012	0.116	0.000282
1999	0.186	0.134	0.038	0.124	0.003825
2000	0.151	0.139	-0.009	0.172	0.000459
2001	0.106	0.129	-0.032	0.130	0.000570
2002	0.141	0.132	0.011	0.097	0.001865
2003	0.146	0.136	0.007	0.143	0.000006
2004	0.155	0.142	0.009	0.144	0.000136
2005	0.158	0.147	0.005	0.151	0.000058
2006	0.177	0.155	0.014	0.152	0.000630
2007	0.166	0.158	-0.002	0.169	0.000013
2008	0.131	0.150	-0.023	0.156	0.000642
2009	0.095	0.134	-0.031	0.127	0.001015
2010	0.129	0.133	0.010	0.103	0.000636