Does Ethnicity Influence Beer Consumption and Brand Choice?

Matthew Aaron Looney

February 15, 2018

Abstract

The beer industry in the United States represents more than $350 billion dollars of total economic impact. This makes the industry one of the largest consumer product sectors in America. As a category of consumer goods, beer is one of the most dynamically changing products in the country. The industry responds quickly to changes in consumer preference and shifting consumer demographics. The continued ability to understand and predict shifting consumer demand is paramount to the industries' ability to continue to innovate and selectively appeal to the heterogeneous consumer. It is well known that consumer demographics play a key role in determining consumer preference. A recent study by Lopez and Matschke (2012) examine consumer preference for beer using market level data. The authors assume a heterogeneous agent model with age and income variation across consumers. Consistent with expectations, they find a strong relationship between increased age and income and lower sensitivity to price. However, the study omits other relevant consumer demographic variables, such as ethnicity. This study seeks to understand more deeply how additional consumer characteristics influence preference for beer; and in particular, how ethnicity influences beer consumption and brand choice. The methodological approach taken in this study is to estimate a differentiated demand for beer using a random coefficients discrete choice model. Information Resources, Inc. (IRI) brand level data will be used as well as consumer level demographic data from the Current Population Survey (CPS).

1 Introduction

According to a study by the Pew Research Center, by the year 2055 the United States will no longer have a single racial or ethnic majority (Cohn and Caumont 2016). In 1965, 84% of Americans were non-Hispanic Whites, 11% Black, 4% Hispanic, less than 1% Asian and less than 1% other. By the year 2015 the ethnic distribution of the United States had shifted with non-Hispanic whites now representing 62% of the U.S. population, 12% Black, 18% Hispanic, 6% Asian and 2% other.[1] Over a 50 year period the American demographic landscape had changed with large increases in Hispanic and Asian representation. These divergent trends are expected to continue long into the future at variable rates depending on how immigration policy morphs with fluctuations in the political landscape (Passel and Rohal 2015).

When discussing ethnic trends it has become important to distinguish ethnicity from race and to define these terms precisely. Unfortunately this is not a trivial task. The classic definitions of race and ethnicity used by sociologists and anthropologist are given by the following two definitions. Race refers to a category of people who share certain inherited physical characteristics, such as skin color, facial features, and stature. Using physical differences as their criteria, scientists at one point identified as many as nine races (Smedley 1998).[2] Ethnicity refers to the shared social, cultural, and historical experiences, stemming from common national or regional backgrounds, that make subgroups of a population different from one another. Within the context of ethnicity, an ethnic group can be defined as a subgroup of a population with a set of shared social, cultural, and historical experiences; with relatively distinctive beliefs,

[1] Whites, Blacks and Asians include only single-race non-Hispanics. Asians include Pacific Islanders. Hispanics are of any race. Source: Pew research Center estimates based on adjusted census data

[2] African, American Indian or Native American, Asian, Australian Aborigine, European (more commonly called "white"), Indian, Melanesian, Micronesian, and Polynesian

values, and behaviors; and with some sense of identity of belonging to the subgroup (Barkan 2016).

There has been much debate and disagreement within the social science community when dealing with issues of race and ethnicity. However, the current general consensus is that biological definitions of race are antiquated and largely irrelevant. Issues of racial classification based on genetic profiling have muddied the waters even further.[3] With the current level of genetic diversity present in the world today it is not possible to classify an individual as purely one race over another. The idea of racial classification based on biology has dissolved and the idea of ethnic identity has emerged as the preferred method of individual demographic identifier. Due to the complexity of the problems related to using biological race as a demographic identifier we will instead use the concept of ethnicity as defined above. An increasing number of data sets are including questions about ethnicity. This self reporting of ethnic identity is fundamentally inline with the widely accepted definitions of ethnicity and thus we feel it more appropriate to use ethnic demographic information when studying consumer preference.

Distinct patterns of immigration from Latin America and Asia is driving much of the current change in the demographic landscape. Hispanics and Asians are estimated to represent 70% of the total immigrant population by the year 2055 with 34% and 36%, respectively (Cohn and Caumont 2016). But immigration patterns represent more than simple statistics. Immigrants are people who bring with them ethnic identity and culture. This ethnic identity is embedded in food, religion, language, and to varying degrees, permeates every aspect of an individuals identity. The United States has always been a nation of immigrants. Since it's founding, the country has

[3]With the availability of DNA testing by companies such as Ancestry.com and 23andMe many people discover the true racial diversity within their genes.

undergone three distinct episodes of immigration. The most recent between 1850 to 1930 when large waves of immigrants from Germany, Italy, Ireland, Greece and many eastern European nations landed on American soil. These immigrants changed the American food scene by bringing their food preferences with them from their home countries. Given the current changes in the ethnic landscape of the United States, it is widely accepted that a fourth episode of immigration is underway. Consumer food preference last generation is looking very different from consumer food preference in the current generation. In fact, at the current pace of ethnic change, shifts in inter-generational food preference are becoming a reality that the U.S. food industry will need to manage to appeal to the heterogeneous consumer. Understanding how food preferences change with large populations from Latin America and Asia settling in the fabric of the United States is an important research question for economist to tackle. Many studies incorporate consumer income and age variations when studying demand for differentiated goods but few explicitly model ethnicity as an added demographic variable. We argue that ethnicity is an important component to consider when studying consumer demand. This study seeks to understand more deeply how additional consumer characteristics influence product preference, and in particular, how ethnicity influences beer consumption and brand choice.

One of the largest consumer product sectors in America is the beer industry. As a whole, this industry represents more than $350 billion dollars of total economic impact.[4] As a category of consumer goods, beer is one of the most dynamically changing products in the country. The industry responds quickly to changes in consumer preference and shifting consumer demographics. The ability of the beer industry to change quickly to appeal to a new consumer group makes the industry an

[4]Based on data from 2016 in a study by John Dunham & Associates.

interesting candidate to study with regard to changing ethnic diversity.

1.1 Literature Review

The theoretical framework for projecting consumer choice onto a set of product characteristics originated with work by Kelvin Lancaster (1966). This idea was extended further by Daniel McFadden who developed the econometric model (1973). This projection allowed for the estimation of a large number of products using a finite number of characteristics. Prior to this line of research the problem of dimensionality plagued empirical industrial organization and rendered many interesting problems intractable.

While dimensionality was a large problem to overcome, price endogeneity is present in almost all supply and demand models and discrete choice modeling is no exception. If we fail to account for this endogeneity then our parameter estimates will be biased and depending on the exact model specification, inconsistent. Steven Berry's 1994 Rand paper provides an elegant method to overcome the endogeneity issue (1994). He developed a method to estimate differentiated demand models in the presence of unobserved product characteristics by inverting the market share equation to obtain an estimate of mean utility. Once the implied mean utility is obtained the estimation can proceed in a similar way as homogeneous goods modeling. While Berry's method allows for the handling of endogenous prices through the use of standard linear instrumental variables his method does not address the well known problem of independence of irrelevant alternatives (iia) thereby severely restricting the cross-price elasticity patterns.[5]

[5]For detailed discussions see Domencich and McFadden (1975), Hausman and Wise (1978), (S. Berry et al. 1995), (Nevo 2000)

The incorporation of consumer heterogeneity in discrete choice modeling of differentiated products explicitly solves the iia problem. In a 1995 Econometrica paper by Steven Berry, James Levinsohn and Ariel Pakes (BLP) (1995) it was shown that individual consumer heterogeneity can be applied to a discrete choice model of differentiated goods in a tractable way to uncover consistent estimations of demand elasticities. The BLP model developed a way to break the iia property which allowed for more realistic substitution patterns across products. The BLP model has become a popular model in industrial organization and has become the model of choice when investigating merger simulations.

It is well known that consumer demographics play a key role in determining consumer preference. A search of the literature related to discrete choice modeling of differentiated goods reveal few studies that incorporate other demographic variables beyond income and age. A study by Lopez and Matschke (2012) use a random coefficients multinomial logit model to examine consumer preference for beer using market level data. The authors assume a heterogeneous agent model with age and income variation across consumers. Consistent with expectations, they find a strong relationship between increased age and income and lower sensitivity to price. However, the authors omit other relevant consumer demographic variables, such as ethnicity. One of the few studies to incorporate an ethnic variable in a random coefficients discrete choice model was performed by Charles Romeo (Romeo 2016). His results show that as the percentage of Hispanics in the population increase, the utility gained from consuming Corona decrease and turn negative with 18 to 22 percent Hispanics. Romeo can not offer any clear explanation of these results but suggests that perhaps with increasing percentages of Hispanics in the market, competition from other Latin American beers are moving consumers away from Corona.

1.2 Conceptual Framework

The demand system is obtained by assuming an individual will maximize their utility by selecting a product from the choice set. The conditional indirect utility of individual i selecting product j in market t is given by $U(x_{jt}, \xi_{jt}, p_{jt}, \zeta_i|\theta)$, where x_{jt} is a K-dimensional (row) vector of observable product characteristics, ξ_{jt} is the unobserved (by the researcher) product characteristic, p_{jt} is the price of product j in market t, ζi are individual characteristics, and θ is a set of unknown parameters to be estimated.

Then indirect utility is given by the following:

$$u_{ijt} = \alpha_i f(\cdot) + \beta_i x_{jt} + \xi_{jt} + \varepsilon_{ijt} \tag{1}$$

where, $i = 1, ..., I_t,\ j = 1, ..., J,\ t = 1, ..., T,\ f(\cdot)$ is a functional form for indirect utility[6], α_i is individual i's marginal utility gained from the parameters of the utility function, β_i is a K-dimensional (column) vector of individual-specific taste coefficients, and ε_{ijt} is a mean zero stochastic error term.

The introduction of consumer heterogeneity is done through modeling individual taste preference (α_i, β_i) based on consumer characteristics, ζ_i. The individual consumer characteristics consist of two parts, observed characteristics, D_i and unobserved characteristics v_i. The observed characteristics can be obtained by either sampling from a joint parametric distribution (e.g., the Current Population Survey) or by using sample data to estimate the mean and standard deviation of the characteristics of interest from a population. For example, assume a log-normal distribution on income

[6]Depending on the research question the functional form can take on any one of a number of utility functions. For example, if wealth effects are thought to be important then a Cobb-Douglas functional may be appropriate, if not then quasi-linear may be chosen, see (Petrin 2002)

and take random draws using the mean and standard deviation from census data. The unobserved characteristics, v_i, can be generated by random sampling from a parametric distribution.

Following nomenclature used by Nevo we get the following:

$$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \Pi D_i + \Sigma v_i, \; v_i \sim P_v^*(v), \; \mathrm{D}_i \sim \hat{P}_D^*(D) \tag{2}$$

where, D_i is observed characteristic, v_i is unobserved characteristics, $P_v^*(\cdot)$ is a parametric distribution, $\hat{P}_D^*(\cdot)$ is either a parametric or non-parametric distribution. Π is a matrix of parameter estimates that measure the variation in taste characteristics across demographics and Σ is a matrix of parameter estimates for the unobserved demographic characteristics.

For completion of the demand system we need to define the outside good. Without the outside good construction all individuals in all markets would be purchasing all products all of the time. For the econometrician to see a change in quantities demanded with a homogeneous price increase across all products (relative to other markets), we need people moving to the margins which means we need to define the option where people can make a zero purchase. We define the mean utility gained from the outside option as U_{i0t}. It is important to note that the outside option is not identified.[7]

Combining equations (1) and (2) we obtain the expression for the conditional indirect utility

[7] It is possible to identify the outside option with further restrictions and normalization, see Nevo (2000) for a more detailed discussion.

$$
\begin{aligned}
u_{ijt} &= \delta_{jt}(x_{jt}, p_{jt}, \xi_{jt}|\theta_1) + \mu_{ijt}(x_{jt}, p_{jt}, v_i, D_i|\theta_2) + \varepsilon_{ijt}, \\
\delta_{jt} &= \beta x_{jt} - \alpha p_{jt} + \xi_{jt}, \\
\mu_{ijt} &= [-p_{jt}, x_{jt}](\Pi D_i + \Sigma v_i)
\end{aligned} \tag{3}
$$

where, δ_{jt} is mean utility, and μ_{ijt} and ε_{ijt} are the mean zero deviations from the mean utility which represent the random coefficient interactions between the observed and unobserved parts of the product characteristics and the demographics. The vectors θ_1 and θ_2 contain the linear and nonlinear parameters of the model with $\theta_1 = (\alpha, \beta)$ and $\theta_2 = (\Pi, \Sigma)$.

The model parameters (θ_1 and θ_2) can be estimated using Hansen's generalized method of moments estimation technique (Hansen 1982) and following the procedures detailed in Berry, et. al. (1995) and Nevo (2000).

1.3 Methodology

The methodological approach taken in this study is to estimate a differentiated demand for beer using a random coefficients discrete choice model. IRI Brand level data will be used as well as consumer level demographic data from the Current Population Survey.

2 Notes:

2.1 Not included in paper

Include information about the beer industries ability to shift focus quickly to appeal to changes in age demographics, ie. Millennial, with the development of the craft beer industry; reference PhD thesis work by Toro and subsequent work that stemmed from his thesis.

need to get more clear about how to use a counter-factual to demonstrate and disentangle the demand shift. Is it beer industry responding to changing ethnic diversity that is causing ethnic groups to change their choice preference (marketing targeted to ethnic groups) or is it changing ethnic diversity that is causing the beer industry to change their products to meet a changing demographic variable. Or does it matter This does not need to be in the introduction but needs to be addressed at some point, maybe

Expand the methodology used with some generic formulation of BLP theory; just enough to give the reader an idea of how I am going to preform the study and what information will be obtained, ie elasticities, etc. Will save the in depth formulation for the theory section. Also need to explain the market and brand choice definitions quickly. Again I will save the in depth descriptions for the data and methodology sections

References

Barkan, S. E. (2016). *Sociology : understanding and changing the social world.* Minneapolis: Minneapolis, MN : University of Minnesota Libraries Publishing.

Berry, S. T. (1994). Estimating Discrete-Choice Models of Product Differentiation. *The RAND Journal of Economics*, *25*(2), 242–262.

Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, *63*(4), 841.

Cohn, D., & Caumont, A. (2016). 10 demographic trends that are shaping the US and the world. *Fact Tank, Pew Research Center, March*, *31*.

Hansen, L. P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, *50*(4), 1029–1054.

Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of political Economy*, *74*(2), 132–157.

Lopez, R. A., & Matschke, X. (2012). Home Bias in US Beer Consumption. *Pacific Economic Review*, *17*(4), 525–534.

McFadden, D., & others. (1973). Conditional logit analysis of qualitative choice behavior.

Nevo, A. (2000). A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand. *Journal of Economics & Management Strategy*, *9*(4), 513–548.

Passel, J., & Rohal, M. (2015). *Modern Immigration Wave Brings 59 Million to US, Driving Population Growth and Change Through 2065.*

Petrin, A. (2002). Quantifying the Benefits of New Products: The Case of the Minivan. *Journal of political Economy*, *110*(4), 705–729.

Romeo, C. J. (2016). Incorporating Prior Information into A GMM Objective For Mixed Logit Demand Systems. *Journal of Industrial Economics*, *64*(2), 336–363.

Smedley, A. (1998). "Race" and the Construction of Human Identity. *American Anthropologist*, *100*(3), 690–702.