Who are Consuming Hemp Products in the U.S.? Evidence from Nielsen Homescan Data

By

GwanSeon Kim
Department of Agricultural Economics
University of Kentucky
(tate.kim@uky.edu)
Phone: 859-218-5210

Tyler Mark
Department of Agricultural Economic
University of Kentucky
(tyler.mark@uky.edu)
Phone: 859-257-7283

Selected Paper prepared for presentation at the Southern Agricultural Economics Association's 2018 Annual Meeting, Jacksonville, Florida, February 3-6 2018

Acknowledgement**:** *"Calculated (or Derived) based on data from The Nielsen Company (US), LLC and marketing databases provided by the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business." and "The conclusions drawn from the Nielsen data are those of the researchers and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein."*

Introduction

Over the last two decades, industrial hemp (also known as hemp) globally has received a great deal of interest in being grown as an agricultural crop. Based on a section 7606 of the 2014 Farm Bill, the industrial hemp is defined as a variety of the *Cannabis sativa* plant species with delta-9 tetrahydrocannabinol concentration (THC) and no more than 0.3 percent on a dry weight basis. Industrial hemp has fifty thousand plus uses that range from fiber to health products, and more than 30 countries currently grow industrial hemp (Johnson, 2017). The Kentucky Department of Agriculture (KDA) reports that approximately 55,700 metric tons of industrial hemp is produced around the world in each year. Approximately 70 percent of industrial hemp in the world are produced in China, Russian, and South Korea.[1]

According to Fortenbery and Bennett (2004), industrial hemp on the production side includes environmental benefits such as low pesticide and herbicide requirements, a wide range of adaptability for agronomic conditions, increased profit centers for U.S. farmers, and relatively low water needs. Other benefits of the industrial hemp on the demand side are increased efficiency compared to other inputs for industrial use, health benefits of both hemp oil and hemp seed consumption, and competitive use in textile manufacturing (Fortenbery and Mick, 2014). Although there are many benefits of industrial hemp, growing industrial hemp in the U.S. is still classified as a Schedule 1 substance. Industrial hemp and marijuana are botanically the same plant species as Cannabis sativa even though they are genetically different from a chemical makeup and cultivation practice standpoint (Cherney and Small, 2016, Datwyler and Weiblen, 2006, Johnson, 2017).

[1] http://www.kyagr.com/marketing/industrial-hemp.html

Since there is no commercial production in the U.S. due to the production of restrictions, all hemp-based products are imported from other countries. For instance, raw and processed hemp fiber is dominantly imported from China whereas hemp seed and oilcake are mostly imported from Canada (Johnson, 2017). Figure 1 provides the total value of U.S. hemp imports from 2010 to 2015, and it shows that total value of imported hemp is increasing.

<Insert Figure 1 Here>

Johnson (2012) claims more than 25,000 hemp products are available in global markets including agriculture, textiles, automotive, furniture, food, personal care, construction, paper, and even recycling. Based on Johnson (2017), Hemp Industries Association (HIA) estimates that annual growth in U.S. hemp retail sales is averaged more than 15% from 2010 to 2015. The author also mentions that the growth is explained by increased sales of hemp-based body products, supplements, and foods by accounting for more than 60% of the value of U.S. retail sales. Recently, Vote Hemp, which is the national, single-issue, nonprofit organization and nation's leading grassroots hemp advocacy organization, estimates the total retail value of hemp products sold in the U.S. in 2016 is approximately $688 million including food and body products, clothing, auto part, building materials, and other products.[2]

Even though there is no commercial hemp production in the U.S., retail sales for hemp production is increasing over time. Based on our best knowledge, no study has been investigated and examined factors that affect consumption of hemp or hemp by-products. In this study, we investigate the important economic and demographic characteristics that are associated with

[2] Vote Hemp is dedicated to the acceptance of and free market for industrial hemp, low-THC oilseed and fiber varieties of Cannabis and working to change state and federal laws to allow commercial hemp farming. More information about estimates of 2016 Annual Retail Sales for Hemp Products are available at http://www.votehemp.com/PR/PDF/4-14-17%20VH%20Hemp%20Market%20Data%202016%20-%20FINAL.pdf

hemp consumption and investigate their effects on expenditure in the U.S. by utilizing Nielsen's consumer panel data from 2008 to 2015. This study employs Heckman selection model since this model provides different parameters of the choice and consumption processes by controlling for non-randomly selected samples. Therefore, we specifically identify the impact of either economics or household characteristics on the probability of purchasing hemp products and which factors impact on the total expenditures on hemp products. Furthermore, this study incorporates information about U.S. state industrial hemp legislation based on the hypothesis that probability of purchasing and expenditures on hemp products are relatively higher in specific states where hemp bills and resolutions are introduced.

Findings from this study will contribute to acceptance of free market for the industrial hemp against existing state or federal laws or policies in the U.S. regarding illegalized commercial hemp production. In addition, this paper provides potential market opportunities by not only understanding consumers but also targeting groups of consumers to increase the market share of the hemp products. The rest of the paper is organized as follows. The next section provides a summary of U.S. hemp history and current U.S. hemp production, while the subsequent section describes the main econometric model. The following section describes the data section especially structure of the data and the variable classification. The next section presents results and discussions, while the final section summarizes main results with limitations and directions for future research.

Background

U.S. Hemp History

The first harvest of hemp was estimated around 8500 years ago (Schultes, 1970) and actively cultivated and domesticated around 4000 and 6000 years ago in China (Kraenzel, et al., 1998, Vavilov and Dorofeev, 1992). In 1545, hemp was initially introduced in the world after Spanish brought the plant to Chile, and hemp became as important crops in Colonial America since New England first grown the plants for fiber source for household spinning and weaving in 1645 (Ehrensing, 1998, Fike, 2016). Cultivation spread to Virginian and Pennsylvania, and a commercial cordage industry with hemp fiber was developed and flourished in 1775 by settlers who brought hemp from Virginia to Kentucky (Fortenbery and Bennett, 2004). In the mid-1800s, hemp was widely grown as the common use of fine and coarse fabrics, twine, and paper in the U.S. (Johnson, 2012). Between 1840 and 1860, especially, the hemp industry was expanded from Kentucky to Missouri and Illinois due to the strong demand for cordage and sailcloth by the U.S. Navy (USDA, 2000). However, the hemp production declined by the end of the 1800s due to the technological innovation and the discovery of alternative inputs for traditionally hemp-based industries (Fortenbery and Mick, 2014). In 1937, U.S. hemp production was effectively prohibited by the passage of the Marijuana Tax Act, which was placed all *Cannabis* culture as a narcotic drug under the control of the U.S. Treasury Department (Fortenbery and Bennett, 2004, Johnson, 2012). During World War II, hemp production was produced again in the U.S. by an emergency program since World War II interrupted supplies of jute and abaca to the U.S. from the tropics, and the production peaked in 1943 and 1944 (Ehrensing, 1998). According to Johnson (2012), the hemp production reached more than 150 million pounds on 146,200 harvested acres in 1943, especially 140.7 million pounds of hemp fiber and 10.7 million pounds of hemp seed; however, the hemp production declined to 3 million pounds on 2,800 harvested acres in 1948. The declined hemp production after the war was due to the re-imposed legal restriction and re-established jute and abaca imports (Fortenbery and Bennett, 2004, USDA, 2000). Even though a small hemp fiber industry continued in Wisconsin until 1958, essentially there has been no U.S. hemp production since then (Dempsey, 1975, Ehrensing, 1998, Fortenbery and Mick, 2014).

Current U.S. Hemp Production

The U.S. Congress replaced the 1937 Marijuana Tax Act with the Comprehensive Drug Abuse Prevention and Control Act in 1970 to distinguish between marijuana and hemp, but U.S. Drug Enforcement Agency (DEA) policy eventually treated marijuana and hemp as the same plant (Cherney and Small, 2016). Even though the domestic hemp production has been restricted due to the federal laws and drug policy in the U.S., there has been an active movement to legalize industrial hemp production in the U.S. for the last two decades (Fortenbery and Mick, 2014). From the mid-1990s, hemp in the U.S. was substantially resurfaced as the potential uses of the plant after Europe and Canada legalized and issued licenses to allow industrial hemp production (Fike, 2016). To the begin with Kentucky in 1994, a total of 19 States have introduced hemp legislation since 1995, and nine States passed legislation authorizing feasibility studies for domestic industrial hemp production in 1999 (USDA, 2000). Even though hemp is still classified as Schedule 1 controlled substance under the Controlled Substances Act (CSA), section 7606 of the U.S. Agricultural Act of 2014 legalized state departments of agriculture and certain research institutions to grow hemp as a pilot program for research purposes (Cherney and Small, 2016, Johnson, 2017). The Vote Hemp reports that 31 states are recorded with enacted hemp laws, and those of 15 states are allowed to grow and cultivate the hemp in 2016. Figure 2 shows the states with both enacted hemp laws and active hemp pilot program. In Figure 2, colored states represent cultivating hemp states including California, Colorado, Hawaii, Indiana, Kentucky, Maine, Minnesota, Nebraska, Nevada, New York, North Dakota, Oregon, Tennessee, Virginia, and West Virginia with planted acres. Compared to other cultivating states, the States of Colorado and Kentucky are especially shown as main cultivating states with higher planted acres: 5,922 and 2,525 acres, respectively. The states with a slanting line represent the legalized hemp states with the enacted hemp law even though there is no cultivation.

<Insert Figure 2 Here>

Conceptual Framework

McFadden (1980) provides the theoretical framework for a qualitative choice model based on a random utility model. The following discussion of the random utility model is based on Alviola and Capps (2010). The product purchase made by the consumer between hemp by-products and conventional products can be modeled as a binary choice. The outcome variable Y_i, therefore, takes a value either 1 if consumers purchase hemp by-products or 0 with conventional products. A utility function with the specification of the binary choice can be defined as following:

$$U(W_i, \varepsilon_i)$$

where utility is a function of the W_i, the covariates in the decision process, $i = 1, 2, \ldots, n$, with n representing the number of households in the sample. The binary choice assuming the existence of the utility function can be rewritten as

$$U_1 = W_1\gamma_1 + e_1$$

$$U_0 = W_0\gamma_0 + e_0$$

where U_1 and U_0 indicate the different level of utility based on the choice with purchasing hemp by-products (U_1) and conventional products (U_0), and e_1 and e_0 are error terms. If the household *i* makes the purchase of hemp by-products, then the probability of choosing hemp by-products can be written as

$$Pr(Y_i = 1) = Pr(U_1 > U_0)$$

$$Pr(Y_i = 1) = Pr(W_1\gamma_1 + e_1 > W_0\gamma_0 + e_0)$$

$$Pr(Y_i = 1) = Pr(e_0 - e_1 < W_1\gamma_1 - W_0\gamma_0)$$

$$Pr(Y_i = 1) = Pr(\mu < W_1\gamma_1 - W_0\gamma_0)$$

where $\mu = (e_1 - e_0)$ is the random variable that follows normal distribution if e_1 and e_0 are normally distributed. With the normal distribution of μ, the probability of purchasing hemp by-products is consequently expressed as

$$Pr(Y_i = 1) = \phi(W_1\gamma_1 - W_0\gamma_0)$$

where ϕ is the cumulative distribution function (cdf). The equation above holds across all households, $i = 1, 2, \dots, n$. Then, ϕ represents the standard normal cumulative distribution function by standardization of μ. In this way, the use of the probit model is justified to investigate the decision on purchasing hemp by-products. In addition, the use of the probit model to investigate the decision on purchasing no hemp by-product is also justified by the binary nature of the choice problem.

Data Description

The main data source for this study is Nielsen Consumer Panel data. The consumer panel data was started since 2004 and updated with a 2-year time lag. The database contains information about the product purchase made by a representative panel of households, approximately 40,000-60,000 households, across all retail channels in all U.S. markets, including food, non-food grocery products, health and beauty aids, and general merchandise. The panelist households continuously provide information, what products they purchase, as well as where and when they make purchases based on the scanned Universal Product Code (UPC) barcode from in-home scanners. Therefore, the Nielsen Consumer Panel data includes detail information about demographic and geographic information of the panelists, products, product characteristics, retail channels, and geographies of major markets.

This study utilizes Nielsen Consumer Panel Data from 2008 to 2015 by focusing on hemp and hemp byproducts. Consumer Panel product data are organized based on hierarchy as follow: departments, product groups, product modules, and UPC codes. First, we employ a searching index function based on a string of characters that include “hemp” to identify the product hierarchy. Table 1 shows a number of observations in different groups of hemp products by the product departments in Nielsen Consumer Panel data from 2008 to 2015. Since most hemp products are found in the product groups of cereal, nuts, vitamins, and medications, this study considers and focuses on only those four product groups.

<Insert Table 1 Here>

Second, we narrow the product groups down to product modules in order to identify whether there are any missing information or irrelevant products to explain those four product groups by looking at next hierarchy, which is product module. Table 2 shows how many observations are included in each product module by the product groups. Based on Table 2, this study excludes the product modules of cereal (hot), ready to eat, and nut (cans) since there are no or only a few observations in the product modules to represent the product groups.

< Insert Table 2 Here>

Third, we collect all households from the Nielsen Consumer Panel data and limit the panelists based on purchases of cereal (granola & nature valley), nut (nuts bag), vitamins (nutritional or protein supplements), and medication (lip remedies). Finally, we exclude households based on uniquely assigned store code in each household by assuming that those households do not have accessibility to buy hemp by-products. Through these steps, we explicitly classify the hemp consumers and estimate the probability of purchasing hemp byproducts and impact of characteristics of households on the total hemp expenditures. Table 3 shows the number of observations for each product with the proportion of hemp products. The nuts, for example, there are total 15,241 households who consume nuts from 2008 to 2015, and those of 11.20 percent households consume hemp nuts.

<Insert Table 3 Here>

The demographic and socioeconomic characteristics, especially, education level, age, racial, and ethnic, in Nielsen's consumer data contains both the male and the female head of households. Since the head of household is either male or female head, this study mainly uses female demographic information by assuming that females make the majority grocery shopping.

This assumption is consistently applied to previous studies such as Dettmann (2008) and Alviola and Capps (2010) that use the Nielsen's consumer data. If the female head of household does not exist, then the head of household is replaced with the male head. Table 4 shows summary statistics of variables used in the analysis. Many of demographic and socioeconomic information in Nielsen's consumer data are classified into many different group categories. This study reclassifies some of them used as explanatory variables. The reclassification of the explanatory variables is as follows. The income in Nielsen is initially classified into 16 different categories, ranging from less than $5,000 to above $200,000. We reclassify 16 income categories into three categories: low if household income is less than $30,000, middle if household income is between $30,000 and $70,000, and high if household income is above $70,000. The age of household head is reclassified from nine categories into three categories: less than 40 years, between 40 and 64 years, and above 64 years. Finally, the education of household head is reclassified from six into four categories: high school or less, some college, college graduate, and post-collegiate.

<Insert Table 4 Here>

Addition to the demographic and socioeconomic characteristics, this study incorporates the variable of hemp state. Since the hemp products are not very popular, and U.S. market is mostly dependent on imports, consumers in the U.S. may have lack of information about the hemp products. In addition, we hypothesize that households in states where hemp bills and resolutions introduced are more likely to be exposed to buy hemp byproducts compared to the households in other states.

Empirical Methodology

This paper employs the Heckman sample selection approach (also called a two-step model) developed by Heckman (1979) to correct the sample selection bias from the non-randomly selected samples. In other words, this study investigates the factors that affect total expenditure for hemp consumers with selection decision whether consumers make purchases on hemp by-products or not by themselves. In addition, we have the expenditure only for the households who buy hemp by-products. The Heckman selection model is different from other approaches such as Tobit model and Cragg's model (also known as hurdle model) for the censored data (i.e., truncated sample) in that the Heckman model is based on incidental truncation rather than truncation. The Heckman approach takes place in two stages.

First Stage of the Heckman Model

The first stage is estimated by the probit model (i.e., selection model) by assuming that error terms are normally distributed. The probit model is defined as follows:

$$Pr(z_i = 1) = \Phi(W_i\gamma)$$

where z_i is an indicator that takes on value of 1 if the household *i* buys hemp by-product and 0 otherwise, Φ is the standard normal cumulative distribution function, and W_i is the vector of explanatory variables for decision to buy hemp by-products. In the first stage, we obtain estimates of γ by Maximum Likelihood Estimation (MLE), and the inverse Mills ratio (IMR) for each household in the selected sample can be estimated as follows:

$$IMR = \hat{\lambda}_i(W_i\hat{\gamma}) = \frac{\phi(W_i\hat{\gamma})}{\Phi(W_i\hat{\gamma})}$$

where $\phi(W_i\hat{\gamma})$ is the estimated probability density function (pdf), and $\Phi(W_i\hat{\gamma})$ is the cdf. The calculated IMR indicates the probability that the household *i* decided to buy hemp by-products over the cumulative probability of the household's decision. In addition, the IMR captures all the effects of the omitted variables (Alviola and Capps, 2010).

Second Stage of the Heckman Model

In the second stage of the Heckman model, we include estimated IMR as an additional explanatory variable to control the endogeneity since the part of the error term for which the decision to buy hemp by-products influence the total expenditure. Therefore, the regression model for the selected sample in the second stage is mathematically formed as

$$E(Y_i|z_i = 1) = X_i\beta + \alpha\hat{\lambda}_i(W_i\hat{\gamma}) + v_i$$

where Y_i represents the total expenditure of hemp by-products by the *i*th household, W is the vector of variables that explain the decision to purchase hemp by-products, X is the vector of explanatory variables associated with the total expenditure of the hemp by-products, and α is the parameter related to the IMR.

Marginal Effects of the Heckman Model

Following discussion about the marginal effects of the Heckman model is based on Saha, et al. (1997) and Alviola and Capps (2010). Let X_{ij} denote the *j*th regression, and it is common for both W_i and X_i. Then estimated marginal effect (ME) of a change in the regressor is defined as

$$\widehat{ME}_{ij} = \frac{\partial E(Y_i|z_i = 1)}{\partial X_i} = \beta_j + \alpha\frac{\partial IMR_i}{\partial X_{ij}}$$

Therefore, the marginal effect of the independent variables on Y_i in the observed sample is composed of two parts. First, there is a direct effect of the expected expenditure on hemp by-products captured by β_j. Second, the indirect effect is captured by a change in the IMR with respect to a unit change in X_{ij}. The equation above can be simplified and rewritten as

$$\widehat{ME}_{ij} = \hat{\beta}_j - \hat{\alpha}\hat{\gamma}\left(W_i\hat{\gamma}\hat{\lambda}_i + (\hat{\lambda}_i)^2\right)$$

where $\widehat{ME}_{ij}$ represents the marginal effect of the *j*th explanatory variable for the *i*th household, $\hat{\beta}_j$ is a parameter estimates for the *j*th explanatory variable in the second stage of the Heckman model, $\hat{\alpha}$ is an estimated parameter for the IMR variable, $\hat{\gamma}$ is an estimated parameter of the *j*th explanatory variable in the first stage of the Heckman model, $W_i\hat{\gamma}$ is the prediction from the probit model for the *i*th household, and $\hat{\lambda}_i$ is an estimated the IMR for the *i*th household who purchase hemp by-products. Saha, et al. (1997) and Alviola and Capps (2010) ague that $\widehat{ME}_{ij} = \hat{\beta}_j$ if and only if $\hat{\alpha} = 0$, and this case is unlikely event that the errors have zero covariance in both first- and second- stage estimation equations; therefore, $\widehat{ME}_{ij} \neq \hat{\beta}_j$ in general. In this paper, we evaluate the marginal effect at the sample mean since the estimated marginal effect is observation dependent as follows:

$$\widehat{ME}_{ij}|_{\boldsymbol{sample\ mean}} = \hat{\beta}_j - \hat{\lambda}\hat{\gamma}_j\left((\bar{W}\hat{\gamma})\bar{\hat{\lambda}} + \bar{\hat{\lambda}}^2\right)$$

where $\bar{W}$ denote the vector of regressor sample mean and $\bar{\hat{\lambda}} = \frac{\phi(\bar{W}\hat{\gamma})}{\Phi(\bar{W}\hat{\gamma})}$ is the IMR evaluated at the means.

Empirical Specification

For the model specification, the first-stage Heckman model, probit model, is hypothesized as a function of the socioeconomic and demographic characteristics including household income, household size, marital status, age, education, race and ethnicity of the household head, and hemp state. The mathematical expression of the probit model for the decision to purchase hemp by-products is written as follows:

$$Pr(z_i = 1) = \gamma_0 + \gamma_1 M\_Income + \gamma_2 H\_Income + \gamma_3 Age2 + \gamma_4 Age3 + \gamma_5 HHSize + \gamma_6 Married + \gamma_7 Edu2 + \gamma_8 Edu3 + \gamma_9 Edu4 + \gamma_{10} White + \gamma_{11} Black + \gamma_{12} Asian + \gamma_{13} Hispanic + \gamma_{14} Employ + \gamma_{16} Hemp\_State + \epsilon_i$$

A description of the variable names in equestion above is based on Table 4 with associated descriptive statistics. Since most of the explanatory variables are either a dummy or indicator variables, the reference categories are excluded in the equation above to avoid the cases that a set of dummy variables is highly correlated with each other (also known as dummy variable trap). The reference categories are reported as baseline in Table 4. Even though we include regional and year dummies in the estimation, we do not report them into the equation above.

The mathematical expression of the second-stage estimation is defined as follow:

$$\ln(Expenditure) = \beta_0 + \beta_1 M\_Income + \beta_2 H\_Income + \beta_3 Age2 + \beta_4 Age3 + \beta_5 HHSize + \beta_6 Married + \beta_7 Edu2 + \beta_8 Edu3 + \beta_9 Edu4 + \beta_{10} White + \beta_{11} Black + \beta_{12} Asian + \beta_{13} Hispanic + \beta_{14} Employ + \lambda_1 IMR + u_i$$

For the dependent variable, this study uses aggregated yearly expenditure, and we transform the dependent variable in logarithm form in order to control heteroskedasticity caused by outliers. In the second stage estimation, we exclude the variable of hemp state in that it is not atypical in hackman selection model. In addition, the variable of *IMR* (*Inverse Mills Ratio*) calculated from the probit model is included to test the selection bias.

Empirical Results

First-Stage Estimation

The results of the first-stage probit model for four different categories are reported in Table 5 including the maximum log-likelihood estimates and McFadden R^2. The marginal effects associated with the estimates of the parameters are also reported in Table 5 since the magnitude of the coefficients does not provide direct interpretation. By looking at the marginal effects in Table 5, households with higher income are more likely to consume cereal and nut, relative to low-income categories with less than $30,000. On the other hands, households with higher age group are less likely consume all food categories except protein compared to the households with age less than 40, indicating young households are more likely to consume the hemp by-products. For the education level, we find most of the categories of hemp by-products except cereal are more likely to be consumed as the level of education increases. In addition, we find that significant regional effects on total expenditure of hemp by-products, but the regional effects vary across the categories of hemp by-products. The states (called “hemp state”) where industrial hemp bills were introduced plays an important role in the purchase of hemp by-products. From Table 5, households in the hemp state are more likely to consume cereal but less likely to consume for nut compared to households in states where there is no hemp legislation.

After the estimation of the probit model, we conduct a prediction success not only to evaluate qualitative choice models but also to assess the usefulness of the probit model as suggested by other studies such as XX, YY, ZZ. Table 6 shows the goodness of fit measures from the probit model for all four categories. As shown in Table 6, the percentage of correct predictions for cereal, nuts, nutrition, and proteins are 76.71%, 88.79%, 97.62%, and 87.27%, respectively.

Second-Stage Estimation

The results of the second stage estimation are reported in Table 6. Within the second stage of results, the lambda (i.e., inverse mills ration) is estimated to test sample selection bias, and it is statistically significant for categories of cereal, nuts, nutrition, and protein at the 0.10, 0.01, 0.10, and 0.01 level, respectively. This indicates the evidence of sample selection bias, and employing hackman selection model is justified. In Table 5, we also reported the marginal effects that are evaluated at the mean. For the second stage estimation, once households made the decision to buy hemp by-products, higher income group of households are positively associated with total expenditure but only in the cereal category. Total expenditure for the categories of nuts and nutrition are positively associated with households in higher age group. Across all categories of hemp by-products, we find higher education level is not statistically related to the total expenditure across most categories. For the different regions, households in South and West consume less for cereal compared to households in East whereas households in Midwest and South consume more nutrition and protein, relative to households in East region. Since hemp-added nutrition and protein are categorized as vitamin products, this finding has two potential implication. First, households in Midwest and South regions perceive hemp byproducts of nutrition and protein are healthier than conventional products. Second, households in East region are less accessible to buy the hemp byproducts of nutrition and protein.

Concluding Remarks

Industrial hemp as a variety of the *Cannabis sativa* plant species has received a great deal of interest in last two decades since there are many benefits in environmental, production, and health. In the global markets, industrial hemp is used in agriculture, textiles, automotive, furniture, food, personal care, construction, paper, and even recycling. In the U.S., retail sales for hemp production is increasing over time even though there is no commercial hemp production due to the production of restrictions. To understand hemp market in the U.S., this study investigates the important sociodemographic factors that are associated with hemp consumption and investigate their effects on total expenditure in the U.S. by utilizing Nielsen's consumer panel data from 2008 to 2015.

By employing Heckman selection model, this study finds that sociodemographic characteristics especially income, age, and education play important role in purchasing and explaining the demand for different categories of hemp by-products. To understand hemp market in the U.S. these finding will provide more targeted marketing strategy. Industrial hemp is still on the Schedule 1 narcotic list and is illegal to produce according to Federal Law. However, there is speculation that it will be removed from this list, given that 30 plus states already have a policy in place for the production of industrial hemp. Based on our best knowledge, there is no empirical study related to the Hemp in the U.S. Thus, findings in this study will begin to fill the knowledge gap on a crop that is increasing consumption and production in the United States. As the industry continues to move forward policymakers are going to need a deeper understanding of the factors driving the industry. Not only will this manuscript contribute to the industrial hemp literature, but it has the potential to generate significant discussion. Little is known about industrial hemp and there are many unknowns about everything from its production to its marketing channels. A basic understanding of consumer profiles is a starting point for these discussions.

Table 1. Number of Observations in Hemp-Related Product Groups by Departments (2008-2015)

Departments	Product Groups	2008	2009	2010	2011	2012	2013	2014	2015	Total
Dry Grocery	Bread	44	37	9	14	9	8	0	0	121
	Breakfast	33	19	27	39	54	14	24	12	222
	Cereal	188	193	192	212	180	152	161	186	1,464
	Condiment	0	0	0	0	0	0	0	1	1
	Nuts	39	26	52	95	200	406	476	416	1,710
	Prepared	3	1	3	1	3	0	0	0	11
	Salad Dressing	3	5	2	0	1	0	0	0	11
	Shortening	0	3	0	0	2	8	16	9	38
	Snacks	0	0	0	0	8	13	4	11	36
	Soft Drinks	0	45	22	29	0	0	0	0	96
Health & Beauty Aids	Vitamins	31	54	38	52	62	119	186	197	739
	Baby Needs	10	7	6	4	8	8	14	11	68
	Cosmetics	0	0	0	0	0	0	1	0	1
	Ethnic Hair	16	8	14	7	3	2	3	2	55
	Hair Care	0	1	0	0	0	0	0	0	1
	Medication	26	25	3	3	99	234	291	257	938
Non-Food Grocery	Household	0	0	0	0	0	1	0	0	1
	Personal	6	7	3	4	3	4	3	1	31
	Total	399	431	371	460	632	969	1,179	1,103	5,544

Table 2. Number of Observations in Hemp-Related Product Modules by Product Groups (2008-2015)

Product Groups	**Product Modules**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**Total**
Cereal	Granola & Nature Valley	188	193	192	212	180	150	144	169	1,428
	Cereal (Hot)	0	0	0	0	0	0	10	13	23
	Ready to Eat	0	0	0	0	0	2	7	4	13
Nuts	Nuts (Bags)	36	26	52	95	200	406	476	416	1,707
	Nuts (Cans)	3	0	0	0	0	0	0	0	3
Vitamins	Nutritional	11	19	14	32	35	78	128	134	451
	Protein	20	35	24	20	27	41	58	63	288
Medication	Lip Remedies	26	25	3	3	99	234	291	257	938
	Total	284	298	285	362	541	911	1,114	1,056	4,851

Table 3. Number of Observations for Each Product with Proportion of Hemp Products

Products	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**Total**
Granola & Nature Valley	727	599	502	559	697	877	1,011	992	5,964
	(25.86)	(32.22)	(38.25)	(37.92)	(25.82)	(17.10)	(14.24)	(17.04)	(23.94)
Nuts (*Bags*)	1,313	1,304	642	975	1,674	2,859	3,395	3,079	15,241
	(2.74)	(1.99)	(8.10)	(9.74)	(11.95)	(14.20)	(14.02)	(13.51)	(11.20)
Nutritional	1,741	1,744	848	1,211	2,194	3,517	3,966	3,657	18,878
	(0.63)	(1.09)	(1.65)	(2.64)	(1.60)	(2.22)	(3.23)	(3.66)	(2.39)
Protein	105	125	101	132	277	457	569	500	2,266
	(19.05)	(28.00)	(23.76)	(15.15)	(9.75)	(8.97)	(10.19)	(12.60)	(12.71)
Lip Remedies	279	250	84	139	284	582	608	579	2,805
	(9.32)	(10.00)	(3.57)	(2.16)	(34.86)	(40.21)	(47.86)	(44.39)	(34.44)
Total	4,165	4,022	2,177	3,016	5,126	8,292	9,549	8,807	45,154
	(6.75)	(7.41)	(13.09)	(12.00)	(10.55)	(10.96)	(11.49)	(11.80)	(10.66)

Notes: the number of observations in table 3 show total number of observation for each product and parenthesis represents a total number of observation for the only hemp related product.

Table 4. Summary Statistics of Variables Used in the Analysis

		Cereal		Nuts		Nutrition		Protein	
Variable	Description	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Hemp Exp	Total yearly expenditure for Hemp by-product in log	1.40	0.48	2.45	0.54	2.44	0.47	2.69	0.39
Hemp	=1 if HH consume Hemp by product	0.24	0.43	0.11	0.32	0.02	0.15	0.13	0.33
Low Income	=1 if HH income is less than $30,000	0.10	0.30	0.14	0.35	0.14	0.35	0.11	0.31
Median Income	=1 if HH income is between $30,000 and $70,000	0.37	0.48	0.39	0.49	0.39	0.49	0.33	0.47
High Income	=1 if HH income is above $70,000	0.53	0.50	0.47	0.50	0.47	0.50	0.56	0.50
Age1	=1 if HH age is less than 40	0.14	0.34	0.09	0.28	0.07	0.26	0.13	0.34
Age2	=1 if HH age is between 40 and 64	0.65	0.48	0.65	0.48	0.63	0.48	0.67	0.47
Age3	=1 if HH age is above 64	0.21	0.41	0.26	0.44	0.29	0.46	0.20	0.40
HH Size	Size of Households	2.52	1.26	2.39	1.19	2.28	1.16	2.45	1.29
Married	=1 if HH married	0.73	0.44	0.71	0.46	0.67	0.47	0.68	0.47
Edu1	=1 if HH education is High School or less	0.15	0.36	0.20	0.40	0.18	0.39	0.13	0.34
Edu2	=1 if HH education is Some College	0.27	0.44	0.30	0.46	0.31	0.46	0.32	0.47
Edu3	=1 if HH education is College Graduate	0.36	0.48	0.34	0.47	0.35	0.48	0.36	0.48
Edu4	=1 if HH education is Post Collegiate	0.22	0.41	0.16	0.36	0.16	0.36	0.18	0.39
White	=1 if HH is White	0.84	0.37	0.81	0.40	0.81	0.40	0.78	0.42
Black	=1 if HH is African American (Black)	0.06	0.23	0.09	0.29	0.09	0.28	0.11	0.31
Asian	=1 if HH is Asian	0.05	0.21	0.05	0.22	0.05	0.23	0.05	0.21
Other Race	=1 if HH is other races	0.06	0.23	0.06	0.23	0.05	0.23	0.07	0.25
Hispanic	=1 if HH is Hispanic	0.08	0.27	0.08	0.27	0.07	0.26	0.08	0.27
Employ	=1 if HH is employed	0.64	0.48	0.57	0.50	0.56	0.50	0.64	0.48
Hemp State	=1 if HH is living in State with Hemp Legislation	0.37	0.48	0.40	0.49	0.40	0.49	0.48	0.50
Midwest	=1 if HH is living Midwest region	0.20	0.44	0.21	0.41	0.15	0.36	0.17	0.38
South	=1 if HH is living South region	0.26	0.44	0.31	0.46	0.31	0.46	0.29	0.45
West	=1 if HH is living West region	0.35	0.48	0.32	0.47	0.41	0.49	0.41	0.49
East	=1 if HH is living East region	0.19	0.39	0.16	0.37	0.13	0.33	0.13	0.33
Observations		5,959		15,233		18,871		2,263	

Notes: S.D represents the standard deviation. HH represents the head of household, and HH is defined as the female head. If a female of the household does not exist, the HH is the male head.

Table 5. First Stage Probit Estimation Result

	Cereal			**Nuts**			**Nutrition**			**Protein**		
Variable	Coef		M.E	Coef		M.E	Coef		M.E	Coef		M.E
M_Income	0.154	**	0.047	0.151	***	0.026	0.041		0.002	0.035		0.007
	(0.069)		(0.021)	(0.047)		(0.008)	(0.065)		(0.003)	(0.128)		(0.025)
H_Income	0.129	*	0.038	0.167	***	0.029	-0.107		-0.005	-0.077		-0.015
	(0.072)		(0.021)	(0.050)		(0.009)	(0.070)		(0.003)	(0.126)		(0.025)
Age2	-0.199	***	-0.061	-0.083		-0.014	-0.298	***	-0.016	0.084		0.016
	(0.054)		(0.017)	(0.051)		(0.009)	(0.066)		(0.004)	(0.107)		(0.020)
Age3	-0.294	***	-0.082	-0.141	**	-0.023	-0.403	***	-0.016	-0.113		-0.021
	(0.069)		(0.018)	(0.059)		(0.009)	(0.078)		(0.003)	(0.148)		(0.026)
HH Size	-0.084	***	-0.025	-0.021		-0.004	0.005		0.0002	-0.030		-0.006
	(0.018)		(0.005)	(0.015)		(0.002)	(0.021)		(0.001)	(0.034)		(0.007)
Married	-0.066		-0.020	-0.157	***	-0.028	0.098	*	0.004	0.063		0.012
	(0.050)		(0.015)	(0.037)		(0.007)	(0.054)		(0.002)	(0.096)		(0.018)
Edu2	-0.053		-0.016	0.121	***	0.021	0.163	***	0.008	0.241	**	0.049
	(0.060)		(0.018)	(0.042)		(0.008)	(0.063)		(0.003)	(0.117)		(0.025)
Edu3	-0.039		-0.012	0.100	**	0.017	0.178	***	0.009	0.237	**	0.048
	(0.059)		(0.017)	(0.042)		(0.007)	(0.064)		(0.003)	(0.119)		(0.025)
Edu4	0.029		0.009	0.184	***	0.034	0.006		0.0003	-0.035		-0.007
	(0.065)		(0.020)	(0.049)		(0.010)	(0.081)		(0.004)	(0.138)		(0.026)
Employed	0.134	***	0.040	0.015		0.003	-0.006		-0.0003	0.159	*	0.030
	(0.043)		(0.013)	(0.031)		(0.005)	(0.045)		(0.002)	(0.086)		(0.016)
White	-0.259	***	-0.082	-0.206	***	-0.038	-0.024		-0.001	0.095		0.018
	(0.084)		(0.028)	(0.061)		(0.012)	(0.098)		(0.005)	(0.158)		(0.029)
Black	-0.013		-0.004	-0.281	***	-0.041	0.150		0.008	0.019		0.004
	(0.114)		(0.034)	(0.076)		(0.009)	(0.114)		(0.007)	(0.195)		(0.038)
Asian	-0.504	***	-0.123	-0.321	***	-0.045	-0.283		-0.010	0.167		0.035
	(0.123)		(0.024)	(0.088)		(0.010)	(0.149)		(0.004)	(0.223)		(0.051)
Hispanic	-0.011		-0.003	0.147	***	0.027	-0.098		-0.004	0.031		0.006
	(0.073)		(0.022)	(0.053)		(0.010)	(0.089)		(0.004)	(0.148)		(0.030)
Midwest	0.159	***	0.049	-0.097	*	-0.016	-0.192	***	-0.008	0.147		0.030
	(0.059)		(0.019)	(0.051)		(0.008)	(0.067)		(0.002)	(0.124)		(0.027)
South	-0.202	***	-0.058	0.260	***	0.047	-0.321	***	-0.014	0.092		0.018
	(0.059)		(0.016)	(0.046)		(0.009)	(0.060)		(0.002)	(0.115)		(0.023)
West	0.268	***	0.082	0.269	***	0.049	-0.307	***	-0.014	-0.279	**	-0.052
	(0.053)		(0.017)	(0.044)		(0.008)	(0.057)		(0.002)	(0.118)		(0.021)
Hemp State	0.086	*	0.026	-0.055	*	-0.009	0.001		0.0001	-0.027		-0.005
	(0.044)		(0.013)	(0.032)		(0.005)	(0.057)		(0.003)	(0.082)		(0.016)
Constant	-0.302	**		-1.940	***		-2.090	***		-1.068	***	
	(0.145)			(0.125)			(0.189)		(0.002)	(0.297)		
Log Likelihood			-3054.221			-5035.573			-2027.504			-812.334
McFadden R^2			0.069			0.058			0.046			0.058
Observations			5,959			15,233			18,871			2,263

Table 6. Goodness of fit measures from Probit model

Categories	McFadden R^2 (Adjusted)	% of Correct Predictions
Cereal	0.061	76.71%
Nuts	0.053	88.79%
Nutrition	0.034	97.62%
Protein	0.028	87.27%

Table 7. Second Stage Estimation Result

	Cereal			**Nuts**			**Nutrition**			**Protein**		
Variable	Coef		M.E	Coef		M.E	Coef		M.E	Coef		M.E
M_Income	0.080	*	0.085	-0.039		0.058	0.102		0.115	0.089		0.073
	(0.044)		(0.044)	(0.053)		(0.044)	(0.070)		(0.066)	(0.095)		(0.058)
H_Income	0.094	**	0.098	0.004		0.106	0.119		0.087	0.056		0.083
	(0.045)		(0.046)	(0.055)		(0.046)	(0.078)		(0.074)	(0.097)		(0.061)
Age2	-0.110	***	-0.117	0.091	*	0.042	0.226	***	0.136	-0.027		-0.073
	(0.040)		(0.040)	(0.053)		(0.042)	(0.071)		(0.060)	(0.079)		(0.048)
Age3	-0.010		-0.021	0.113	*	0.025	0.206	**	0.083	-0.247	**	-0.192
	(0.049)		(0.048)	(0.066)		(0.054)	(0.091)		(0.075)	(0.106)		(0.063)
HH Size	-0.064	***	-0.067	0.028	*	0.014	-0.035		-0.034	-0.013		0.001
	(0.014)		(0.014)	(0.015)		(0.012)	(0.023)		(0.022)	(0.022)		(0.015)
Married	0.096	***	0.094	0.049		-0.055	0.077		0.107	-0.036		-0.063
	(0.033)		(0.033)	(0.043)		(0.036)	(0.060)		(0.058)	(0.071)		(0.045)
Edu2	0.063		0.061	-0.072		0.008	-0.072		-0.023	0.077		-0.023
	(0.042)		(0.042)	(0.048)		(0.038)	(0.073)		(0.066)	(0.086)		(0.055)
Edu3	0.003		0.002	-0.070		-0.007	-0.031		0.023	0.062		-0.041
	(0.040)		(0.040)	(0.046)		(0.037)	(0.071)		(0.065)	(0.089)		(0.057)
Edu4	-0.029		-0.028	-0.117	**	0.006	0.002		0.004	-0.114		-0.064
	(0.044)		(0.043)	(0.056)		(0.044)	(0.085)		(0.081)	(0.102)		(0.065)
Employed	0.072	***	0.077	-0.061		-0.052	-0.042		-0.044	0.093		0.002
	(0.027)		(0.027)	(0.037)		(0.030)	(0.050)		(0.048)	(0.064)		(0.039)
White	-0.071		-0.081	0.158	**	0.039	0.060		0.053	0.038		-0.034
	(0.055)		(0.055)	(0.064)		(0.052)	(0.090)		(0.088)	(0.113)		(0.056)
Black	-0.119	*	-0.119	0.164	**	0.000	0.122		0.167	0.029		-0.010
	(0.065)		(0.065)	(0.083)		(0.067)	(0.112)		(0.111)	(0.144)		(0.085)
Asian	-0.135		-0.154	0.186	**	-0.014	0.181		0.096	-0.091		-0.179
	(0.084)		(0.083)	(0.092)		(0.073)	(0.175)		(0.166)	(0.161)		(0.098)
Hispanic	-0.144	***	-0.145	-0.040		0.056	0.183	*	0.153	-0.014		-0.016
	(0.049)		(0.049)	(0.060)		(0.049)	(0.099)		(0.095)	(0.107)		(0.056)
Midwest	0.029		0.035	0.064		-0.001	0.254	***	0.195	0.237	***	0.148
	(0.045)		(0.044)	(0.060)		(0.049)	(0.073)		(0.074)	(0.088)		(0.052)
South	-0.105	**	-0.113	-0.083		0.073	0.328	***	0.229	0.247	***	0.173
	(0.041)		(0.041)	(0.053)		(0.042)	(0.073)		(0.071)	(0.081)		(0.049)
West	-0.076	*	-0.066	-0.054		0.115	0.391	***	0.297	0.067		0.184
	(0.039)		(0.038)	(0.052)		(0.042)	(0.074)		(0.071)	(0.092)		(0.059)
Lambda	-0.046	*	–	-0.765	***	–	-0.340	*	–	0.657	***	–
	(0.026)		–	(0.047)		–	(0.194)		–	(0.059)		–
Constant	1.493	***	–	3.698	***	–	3.101	***	–	1.811		–
	(0.097)		–	(0.195)		–	(0.530)		–	(0.209)		–
Log Likelihood			-3,976.466			-6,340.948			-2,278.849			-894.196
Censored			4,532			13,526			18,421			1,975
Uncensored			1,427			1,707			450			288
Observations			5,959			15,233			18,871			2,263

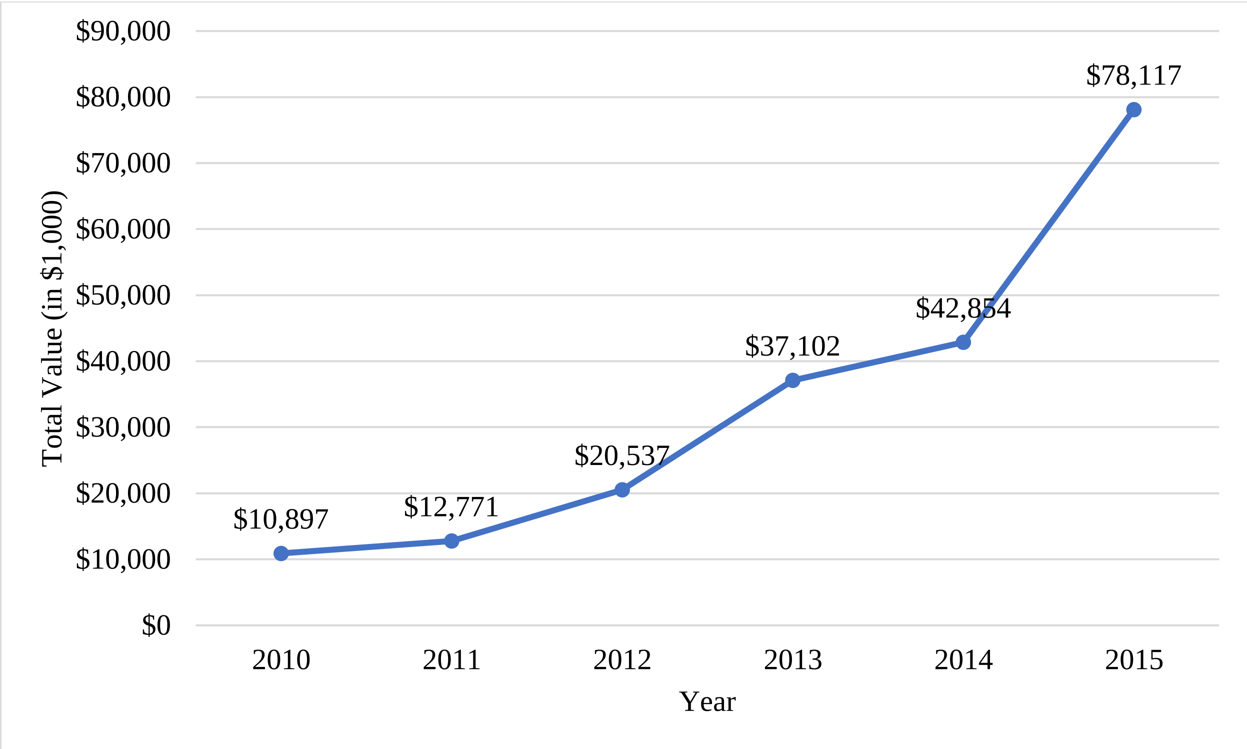

Figure 1. Total Value of U.S. Hemp Imports, 2010-2015

Notes: Main source of total value for hemp imports is obtained from U.S. International Trade Commission, and total hemp imports include hemp seed, hemp oil and fractions, hemp seed oilcake and solids, and true hemp.

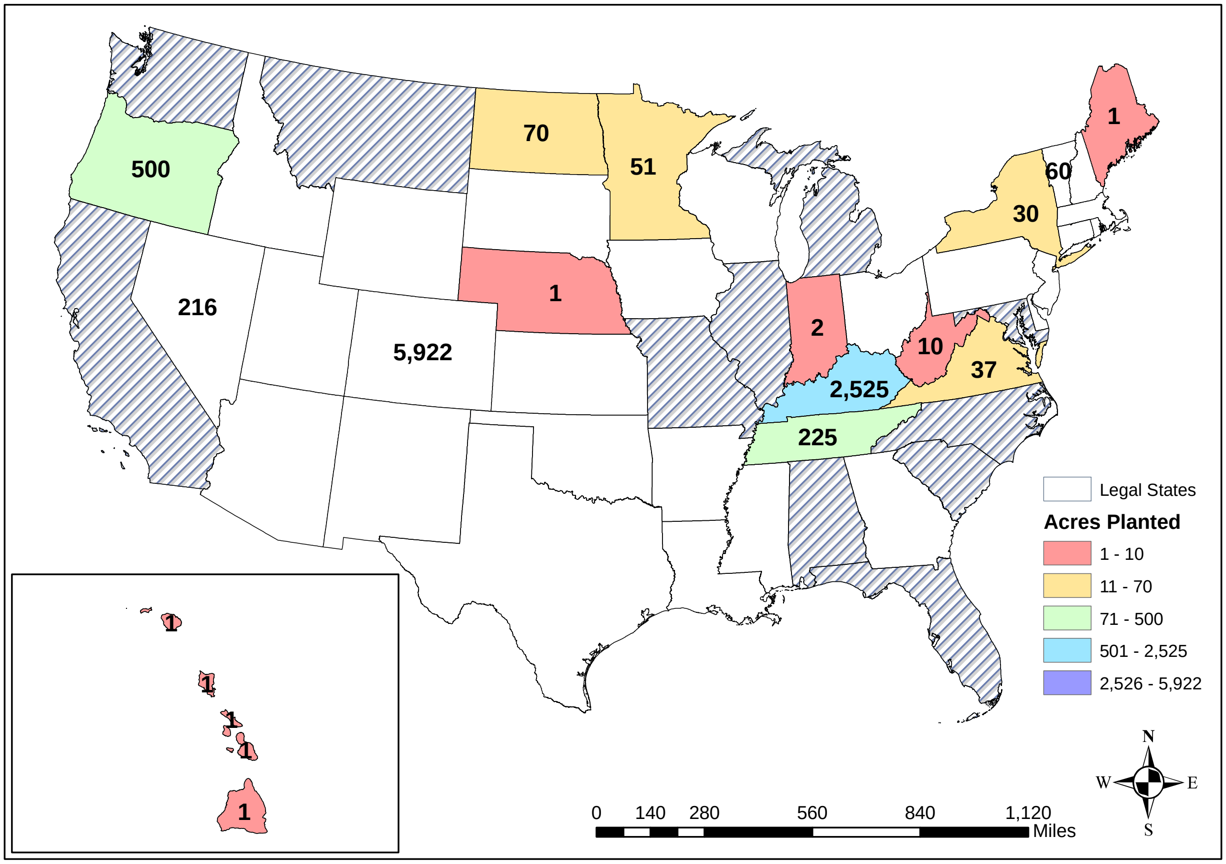

Figure 2. Legalized Hemp States with Acres Planted in 2016

Source: Vote Hemp at http://www.votehemp.com/lobbying/2017-Fly-In-States-Update.pdf

References

Alviola, P.A., and O. Capps. 2010. "Household demand analysis of organic and conventional fluid milk in the United States based on the 2004 Nielsen Homescan panel." *Agribusiness* 26:369-388.

Cherney, J.H., and E. Small. 2016. "Industrial hemp in North America: Production, politics and potential." *Agronomy* 6:58.

Datwyler, S.L., and G.D. Weiblen. 2006. "Genetic variation in hemp and marijuana (Cannabis sativa L.) according to amplified fragment length polymorphisms." *Journal of Forensic Sciences* 51:371-375.

Dempsey, J.M. 1975. *Fiber crops*: Univ. Presses of Florida.

Dettmann, R.L. (2008) "Organic produce: Who's eating it? A demographic profile of organic produce consumers." In *American Agricultural Economics Association Annual Meeting, Orlando.* pp. 27-29.

Ehrensing, D.T. "Feasibility of industrial hemp production in the United States Pacific Northwest." Corvallis, Or.: Agricultural Experiment Station, Oregon State University.

Fike, J. 2016. "Industrial Hemp: Renewed Opportunities for an Ancient Crop." *Critical Reviews in Plant Sciences* 35:406-424.

Fortenbery, T.R., and M. Bennett. 2004. "Opportunities for commercial hemp production." *Review of agricultural economics* 26:97-117.

Fortenbery, T.R., and B.T. Mick. 2014. "Industrial Hemp: Opportunities and Challenges for Washington."

Heckman, J. 1979. "Sample specification bias as a selection error." *Econometrica* 47:153-162.

Johnson, R. (2017) "Defining "Industrial Hemp": A Fact Sheet." In., LIBRARY OF CONGRESS WASHINGTON DC CONGRESSIONAL RESEARCH SERVICE.

--- (2017) "Hemp as an agricultural commodity." In., LIBRARY OF CONGRESS WASHINGTON DC CONGRESSIONAL RESEARCH SERVICE.

--- (2012) "Hemp as an agricultural commodity." In., LIBRARY OF CONGRESS WASHINGTON DC CONGRESSIONAL RESEARCH SERVICE.

Kraenzel, D.G., et al. 1998. "Industrial hemp as an alternative crop in North Dakota." *Agricultural Economics Report* 402.

McFadden, D. 1980. "Econometric models for probabilistic choice among products." *Journal of Business*:S13-S29.

Saha, A., O. Capps, and P.J. Byrne. 1997. "Calculating marginal effects in dichotomous-continuous models." *Applied Economics Letters* 4:181-185.

Schultes, R.E. 1970. *Random thoughts and queries on the botany of cannabis*: J. & A. Churchill.

USDA, E. 2000. "Industrial Hemp in the United States: Status and Market Potential."

Vavilov, N.I., and V.F. Dorofeev. 1992. *Origin and geography of cultivated plants*: Cambridge University Press.