2019

UTILIZING LARGE SCALE DATASETS TO EVALUATE ASPECTS OF A SUSTAINABLE BIOECONOMY

GwanSeon Kim
University of Kentucky, tate.kim@uky.edu
Digital Object Identifier: https://doi.org/10.13023/etd.2019.272

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

GwanSeon Kim, Student

Dr. Tyler B. Mark, Major Professor

Dr. Tyler B. Mark, Director of Graduate Studies

UTILIZING LARGE SCALE DATASETS TO EVALUATE ASPECTS OF A SUSTAINABLE BIOECONOMY

DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Agriculture, Food and Environment
at the University of Kentucky

By

GwanSeon Kim

Lexington, Kentucky

Co- Directors: Dr. Tyler B. Mark, Assistant Professor of Agricultural Economics

and Dr. Michael R. Reed, Professor of Agricultural Economics

Lexington, Kentucky

2019

ABSTRACT OF DISSERTATION

UTILIZING LARGE SCALE DATASETS TO EVALUATE ASPECTS OF A SUSTAINABLE BIOECONOMY

This dissertation combines large scale datasets to evaluate crop prediction, land values, and consumption of a crop being considered to advance a sustainable bioeconomy. In chapter 2, we propose a novel application of the multinomial logit (MNL) model to estimate the conditional transition probabilities of crop choice for the state of Kentucky. Utilizing the recovered transition probabilities the forecast distributions of total acreages for alfalfa, corn, soybeans, tobacco, and wheat produced in the state from 2010 to 2015 can be recovered. The Cropland Data Layer is merged with the Common Land Unit dataset to allow for the identification of crop choice at the field level. Our findings show there are higher probabilities of planting soybeans or wheat after corn relative to corn after corn, tobacco, or alfalfa. In addition, the transition probability of the crop rotation demonstrates that corn will be planted after soybean, and vice versa and that alfalfa has a lower probability of being rotated with other crops from year to year. These findings are expected with traditional crop rotation in the U.S., and a characteristic of a perennial crop, especially for alfalfa. Finally, forecasting results indicate that there are significantly wider distributions in corn and soybean, whereas there is a little variation in the tobacco, wheat and alfalfa acres in the simulation.

In chapter 3, we identify critical consumer-demographic characteristics that are associated with the consumption of products containing hemp and investigate their effect on total expenditure in the U.S. To estimate the likelihood of market participation and consumption level, the Heckman selection model, is employed using the maximum likelihood estimation procedure utilizing Nielsen consumer panel data from 2008 to 2015. Results indicate marketing strategies targeting consumers with higher education and income levels can attract new customers and increase sales from current consumers for this burgeoning market. Head-of-household age in different regions shows mixed effects on decisions to purchase hemp products and consumption levels. Findings will provide a basic understanding of a consumer profile and overall hemp market that has had double-digit growth over the last six years. As the industry continues to move forward, policymakers are going to need a deeper understanding of the factors driving the industry if they are going to create regulations that support the development of the industry.

In chapter 4, we investigate the factors that affect agricultural land values by proposing a new rich dataset, Zillow Transaction and Assessment Data (ZTRAX) provided by Zillow from 2009 to 2014. we also examine whether National Commodity Crop Productivity Index (NCCPI) could be a good indicator of land values or not by comparing two different regression models between county-level cash rent and parcel-level NCCPI. Finally, this study incorporates flexible functional forms of the parcel size to test the parcel size and land values relations. Findings show that factors influencing agricultural land values in states with heterogeneous agricultural lands such as Kentucky are not different from other states with relatively homogeneous agricultural lands. This study also provides suggestive evidence that there is a non-linear relationship between parcel size and land values. Furthermore, we find that a disaggregated NCCPI at parcel-level could be considered an acceptable indicator to estimate agricultural values compared to an aggregated cash rent at county-level.

KEYWORDS: Cropland Data Layer, Common Land Unit, Nielsen Consumer Panel, Industrial Hemp, Zillow Transaction and Assessment Data, Kentucky

GwanSeon Kim

(Name of Student)

07/08/2019

Date

UTILIZING LARGE SCALE DATASETS TO EVALUATE ASPECTS OF A SUSTAINABLE BIOECONOMY

By

GwanSeon Kim

Dr. Tyler B. Mark

Co-Director of Dissertation

Dr. Michael R. Reed

Co-Director of Dissertation

Dr. Tyler B. Mark

Director of Graduate Studies

07/08/2019

Date

Dedicated to my family, Rosa, Aaron, and Anna and our parents whose love and support made this possible

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor Dr. Tyler Mark as an advisor, a dissertation Co-Chair, and a friend. I especially thank Dr. Tyler Mark for providing a great opportunity that my dissertation is supported by the National Science Foundation (award #1355438). His suggestions and insightful comments have been crucial to make me a strong and productive researcher and teacher. He has been and will be continuously a great mentor for my entire life.

I am grateful to the Dissertation Committee members: Dr. Mike Reed, Dr. Leigh Maynard, Dr. Steven Buck, and Dr. Seth DeBolt for providing great suggestions and feedbacks to improve my researches. This dissertation would not be possible without support from staff, faculty, and colleagues in the Department of Agricultural Economics at the University of Kentucky. I appreciate their advice and guidance to complete my Ph.D. study and my graduate career.

Finally, I would like to acknowledge the infinite supports from my family. To my wife, Rosa Hwang, thank you for your endless support and thank you for your patience to endure the long and hard time. To my lovely and precious son, Aaron and daughter, Anna, thank you for giving me the happiness and the strength to overcome the hard times. I especially thank you to my parents in South Korea. The process would not have completed successfully without their support and prayer. Although they were far away, they were always with me in my long journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

The international shift toward green manufacturing and renewable products from biomass has resulted in the concept of the Bioeconomy, which links to energy, agriculture, manufacturing, environmental, and health sectors. Based on the National Bioeconomy Blueprint issued by the White House in April 2012, the bioeconomy is large and rapidly growing segment of the world economy.[1] Based on Oborne (2010), the bioeconomy is defined as economic activities that are associated with the invention, development, production, and use of biological products and processes. The major benefits of the bioeconomy make socioeconomic contributions in both Organization Economics Cooperation Development (OECD) and non-OECD countries. These contributions include improving health outcomes, boosting the productivity of agriculture and industrial processes, and enhancing environmental sustainability.

According to the United State Department of Agriculture (USDA, 2013) and Golden et al. (2013), the bioeconomy contributes to both the overall economy, and our community.[2] From the economic point of view, it contributes approximately $369 billion to the U.S. economy in a single year. This includes 4 million jobs to bio-based industries and $126 billion in sales of bio-based products in 2013. In addition to the economic impacts, the bioeconomy also provide the following benefits to the community: replaced 300 million gallons of petroleum per year, reduced greenhouse gas emission, and 2,250 USDA certified products on the market (Golden et al., 2013).

---

[1] See https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/national_bioeconomy_blueprint_exec_sum_april_2012.pdf

[2] See https://www.biopreferred.gov/BPResources/files/BP_InfoGraphic.pdf

In the U.S., seven major sectors of the biobased product industries that contribute to the U.S. economy are as follows: Agriculture and Forestry, Biorefining, Bio-based Chemicals, Enzymes, Bioplastic bottles and packaging, Forest Products, and Textiles (Golden et al., 2013). Bio-based products, which are generally derived from many different biomass feedstocks, can be categorized by two different products as the first generation (if edible) and second generation (if not edible). The bio-based products such as corn, soybean, and wheat are considered as the first generation of the primary agricultural feedstocks whereas products such as corn stover, miscanthus, and switchgrass are represented as the second generation feedstocks. From an agricultural standpoint, the identification of viable feedstocks and locations that these feedstocks can be produced is one of the first steps in the development of the bioeconomy.

Kentucky, for example, has long been known for its ability to produce forages for the livestock industry. Furthermore, its climate makes it an ideal location for a wide variety of potential feedstocks include but are not limited to switchgrass, miscanthus, sweet sorghum, hemp, kenaf, and corn stover. Many of those biomass feedstocks in development around the nation are not grown or have not been adopted as major crops, and crop producers are unfamiliar with practices and markets associated with them. Considering falling commodity prices in recent years, alternative biomass feedstocks provide producers with at least the same profit per acre as the current commodities being produced. The first essay (Chapter 2) entitled "*Recovering forecast distributions of Crop Composition: Method and Application to Kentucky Agriculture*" estimates the transition probabilities for the five primary row crops produced in Kentucky by employing Crop Data Layer (CDL). In addition, using transition probability from the first Markov chain and simulation technique, we generate probability distributions for each crop and forecast the acreage distribution of each crop to be planted. Findings will contribute to developing strategies to help with the development of the bioeconomy in Kentucky.

Industrial Hemp (also known as hemp) is one of the biomass crops, and there are many different hemp-based products in the U.S. According to Hemp Industries Association and Vote Hemp, total sales in hemp-based products are $573 million in 2015 and $688 million in 2016.[3] In addition, the hemp-based products in the U.S. are categorized by food, Hemp CBD, supplements, personal care, consumer textiles, industrial applications, and other consumer products. Especially for the food sector, the hemp-based food is considered as "*Super Food*" in that it provides several health benefits. All hemp foods are essentially made from hemp seeds, which are known as a nutritionally complete food source in the world due to the perfect balance of omega 3 to omega 6, plus iron, vitamin E, and the essential amino acids.[4] Furthermore, hemp seeds contribute to weight loss, normalize blood sugar levels, improved immune health, improved cholesterol levels, and high protein.[5]

Although demand for hemp-based products and its sales are increasing in the U.S., there is no large-scale commercial production in the U.S. All hemp-based products rely on imports largely from Canada and China. The second essay (Chapter 3) entitled "*Hemp, Hemp, Consumption in the U.S.*" investigates the important socioeconomic and demographic characteristics associated with hemp-based food consumption and their impacts on expenditures in the U.S. by utilizing Nielsen's consumer panel data from 2008

---

[3] See https://www.thehia.org/HIAhemppressreleases/4010402 and http://www.votehemp.com/PR/PDF/4-14-17%20VH%20Hemp%20Market%20Data%202016%20-%20FINAL.pdf
[4] See https://www.leafly.com/news/food-travel-sex/why-are-hemp-seeds-considered-a-superfood
[5] See https://www.nateralife.com/blog/lifestyle/why-is-hemp-a-superfood/

to 2015.[6] Findings in this study will begin to fill the knowledge gap on a crop that is increasing consumption and production in the U.S. As the industry continues to move forward, findings in this study may also open the door to develop a business and marketing plan that creates goals and strategies for marketers, retailers, and other stakeholders. In addition, the findings will provide potential market opportunities by not only understanding consumers but also segmenting groups of consumers to increase the market share of the hemp products.

In the U.S., the average farm real estate value in both nominal and real terms is increasing over time (USDA, 2016). The farmland value is measured based on the productivity (and the returns) of its lands from agricultural activity (Featherstone and Baker, 1987). According to Barnard (2000) and Flanders et al. (2004), the market value of farmland is higher than its use value in agricultural production across the U.S. The portion of the market value, especially from the agricultural production, can be referred as its agricultural use value (Borchers et al., 2014). U.S. agriculture is experiencing fundamental change due to the development of the bioeconomy. The bioeconomy is not only closely related but also significantly affected by agricultural land use and value due to increasing demand and supply of the biomass crop production. Since agricultural land is limited, farmers must compete for land to produce biomass crops. Therefore, it is important to consider agricultural land values and their influence on investment in biomass crops

---

[6] The author(s) would like to thank the Marketing Data Center at the University of Chicago Booth School of Business. Information on accessing this data can be found at http://research.chicagobooth.edu/nielsen/. Results are 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. 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.

productions. The third essay entitled "*Factors affecting Heterogeneous Agricultural Land: The Case of Kentucky*" investigates the factors that affect agricultural land values using a new rich dataset, Zillow Transaction and Assessment Data (ZTRAX) provided by Zillow from 2009 to 2014.[7] This study focuses only on Kentucky by hypothesizing that the factors influencing the farmland values may not be consistent with the findings in the previous studies on farmland values. We then incorporate the flexible, functional forms of the parcel size into the Hedonic framework to calculate specific threshold points where the direction and effect of parcel sizes change. This study strives to make three contributions. First, our study is the first to employ a new rich dataset, ZTRAX, to investigate agricultural land values in a relatively heterogeneous agricultural state. Second, this study will provide suggestive evidence on whether National Commodity Crop Productivity Index (NCCPI) can be used a good indicator of land values rather than cash rent. If NCCPI is the good indicator and substitutable for the cash rent, then this finding will imply that price information should not necessary to be accounted for analyzing the agricultural values. Finally, our results will help land-owners make decisions managing land more efficiently by providing an advanced understanding of the size-effect on land values.

Chapter 5 summarizes the collective findings and provides some discussion of the implications of each chapter of the dissertation.

[7] Data provided by Zillow through the Zillow Transaction and Assessment Dataset (ZTRAX). More information on accessing the data can be found at http://www/zillow.com/ztraz. The results and opinions are those of the author(s) and do not reflect the position of Zillow Group

CHAPTER 2. RECOVERING FORECASTING DISTRIBUTIONS OF CROP COMPOSITION: METHOD AND APPLICATION TO KENTUCKY AGRICULTURE

2.1 Abstract

This paper proposes a novel application of the multinomial logit (MNL) model to estimate the conditional transition probabilities and generate the forecast distributions of total acreages for the five largest crops produced in the state of Kentucky. The transition probability of the crop rotation demonstrates that corn will be planted after soybean (and vice versa) and that alfalfa has a lower probability of being rotated with other crops from year to year. Forecasting results indicate that there are significantly wider distributions in corn and soybean, whereas there is little variation in the tobacco, wheat and alfalfa acres in the simulation.

2.2 Introduction

Improvements in crop production forecasts, yield forecasts, and forecasting methods have been a focus of agricultural economics research for decades. It is due to that fact that crops are traded and priced on commodity exchanges that operates every day. Traders need up-to-date information to make decisions on buying and selling. According to Hayes and Decker (1996), crop production assessments also provide important implications for agribusiness and food management, implying crop production and yield predictions directly influence year-to-year local, state, regional, national and international economies. On a macro level, understanding the determinants of crop acreage and yield forecast helps with the identification and management of the demand and supply of crop production (de Barros Dias, 2017). For example, merchandizers rely on crop supply and demand estimates prepared by both public and private organizations (Vogel and Bange, 1999). Also, estimates and forecasts of acreage and yield can have a significant impact on the futures prices and market volatility (Good and Irwin, 2011), as well as market participants (Egelkraut, et al., 2003). Thus impacting farm income and investment in agriculture, so they are highly anticipated market events (de Barros Dias, 2017, Good and Irwin, 2011).

The National Agricultural Statistics Service (NASS) of the U.S. Department of Agriculture (USDA) is the primary provider of public information on potential crop size and forecast of the average yield and production (Good and Irwin, 2011).[8] In addition to NASS, private companies such as Conrad Leslie and Informa Economics (previously Sparks Companies) develop and use crop forecasts, especially for corn, soybeans, and wheat. (Egelkraut, et al., 2003). The crop production forecasts, as well as the acreage estimation of the NASS, are based on survey data. To be specific, the acreage estimation is reported based on the March and June Agricultural Surveys. In the March Agricultural Survey in 2015, for example, approximately 84,000 farm operators were contacted by mail, internet, telephone, or personal interview, whereas approximately 70,000 farm operators were surveyed in the June Agricultural Survey (Good and Irwin, 2015). Since the survey used to estimate acreage is based on a random sample of farm operators instead of all operators, Good and Irwin (2011) argue that the estimated acreage is subject to sampling error; therefore, it may produce different results with a different sample and inaccurately reflect the population of farm operators.

Even though NASS provides a detailed description of the crop estimating and forecasting, market participants still lack understanding of how acreage, yield, and production forecasts are conducted. This lack of understanding results in a lack of trust in

[8] According to Adjemian and Smith (2012), the USDA also releases the World Agricultural Supply and Demand Estimates (WASDE) at the beginning of May in each year, and it provides forecasts for several crops of annual U.S. production. Thereafter, USDA releases a new WASDE report each month by adding detailed farm surveys, whether forecasts, and expected market development from the NASS and Interagency Commodity Estimates Committees (ICEC).

the objectives of the forecasts (Good and Irwin, 2011). In addition to the sampling error issue with NASS forecasts, the NASS forecasts of acreage only provide mean values of acreage forecasts by county. Therefore, it does not reflect uncertainty in the forecasts related to the risk of decision makers. By shifting to forecasting acreage distributions instead of mean acreage, uncertainty in the forecast values, such as minimum and maximum forecasted acreages, are incorporated. Forecasted distributions can then be used to assess crop sales and total acreage. Savage (2011), for example, utilizes Markov transition probabilities to simulates the distribution of total acreages by crop type whether the forecasted acreage for each crop meet the minimum acreage requirements by Endangered Species Act (ESA) for a variety of endangered species.

The main objectives of this paper are two-fold. First, we estimate the conditional transition probabilities for crop choice utilizing a novel application of the multinomial logit (MNL) model. Secondly, we simulate the distribution of total acreages by crop using the recovered transition probabilities from the first-order Markovian process. Through this process, both sampling error and mean forecast issues can be overcome by using all the fields instead of a random sample of the fields; forecasting the distribution of acreage in addition to the mean acreage. The method proposed in this paper utilizes forecasted distributions for the five largest crops produced in the state of Kentucky. The method, therefore, could make use of publicly available data and provide an additional option for governmental and non-governmental groups trying to predict crop yield, acreage, price forecasts, land values, etc. In addition, response rates on NASS crop acreage and production surveys have been declining since the early 1990s, and it consequently could result in declining the statistical reliability of estimates and forecasts (Johansson, et al., 2017; Schnepf, 2017). Therefore, the method proposed in this study could supplement NASS survey and improve the quality of NASS crop acreage and production estimates.

The remainder of the paper is structured as follows: section two discusses related literature, section three describes and presents data. Section four explains the empirical models including the multinomial logit model and a first-order Markov chain approach. Section five discusses the analysis and presents the results. We conclude the paper in section six and offer policy implications and areas for future research.

2.3 Literature Review

This paper has relevance to the existing literature in two particular areas: (i) importance of evaluating crop rotation and transition probabilities; (ii) literature that uses MNL and first-order Markov process theory in agricultural related studies.

2.3.1 Crop Rotation and Crop Transition Probabilities

Crop rotation (also called polyculture) is defined as growing a series of multiple crops in the same field in alternating years whereas monoculture is defined as growing a single crop in consecutive years in the same field (Martinez and Maier, 2014). Farmers commonly practice crop rotation because its advantages offset its disadvantages. Crop rotation benefits include increased yield (Leteinturier, et al., 2006, Porter, et al., 1997), improved soil fertility (Hendricks, et al., 2014b, Karlen, et al., 2006, Plourde, et al., 2013), reduced greenhouse gas emissions (Halvorson, et al., 2008), and reduced economic risk by having more than one crop as a potential income source (Martinez and Maier, 2014). The traditional crop rotation in western Kentucky is either a Corn-Soybean Rotation or Wheat-Double Crop Soybeans-Corn. In areas with tobacco production, it is typically Tobacco-Tobacco-Alfalfa. Both corn and tobacco require significant nitrogen fertilizer for growth, so that is why they are rotated with either soybeans or alfalfa. Furthermore, incorporating leguminous crops, commonly known as nitrogen-fixing crops such as soybeans, into a rotational sequence with the region's dominant crop will result in increasing the robustness and resilience of local agricultural system (Burgess, et al., 2012, Long, et al., 2014).

2.3.2 The Multinomial Logit Model and Markov Chain Process

The MNL model has been widely used in agricultural related studies especially for modeling land use (Carrión-Flores, et al., 2009, Hardie and Parks, 1997, Lichtenberg, 1989, Plantinga, et al., 1999, Wu, et al., 2004). Lichtenberg (1989) estimates county-level cropland allocation based on the seven major crops in western Nebraska from 1966-1980. Hardie and Parks (1997) employ the MNL model into the land use decision by incorporating heterogeneous land quality in the southeastern U.S. They argue that the MNL model allows errors not only from the use of county averages but also from the use of sample estimates of land use acreage. Plantinga, et al. (1999) simulate carbon sequestration based on estimates of land use share in Maine, South Carolina, and Wisconsin by utilizing the MNL model. Wu, et al. (2004) predict crop choice and tillage practices to assess the economic and environmental consequences of agricultural land-use changes by using the MNL model. Carrión-Flores, et al. (2009) use the MNL model by incorporating spatial dependence in Medina County, Ohio for the determinants of land use choices. Paton, et al. (2014) investigate the impact of rainfall and crop profit margin on crop choice by using MNL regression to generate the crop choice transition probabilities.

Matis, et al. (1985) propose a methodology to forecast crop yields and provide forecast distributions of crop yield by using Markov chain theory. They find that forecasting crop yield distributions are more informative compared to forecasting mean yields. Regarding preciseness, they also find that estimates using the Markov chain and regression approaches are approximately similar. The Markov chain approach, as a non-parametric method, provides point estimates that are not constrained by distribution assumptions whereas the point estimates from the regression approach, a parametric method, depends upon the normal distribution assumption. In this study, we use the first order Markov Chain approach to predict crop acreage distribution by accounting for the dynamics in the analysis by specifying the year-to-year transitions between five crops. Furthermore, the first order Markov chain is appropriate in this study because crop rotation generates dynamic complementarity in crop production. Specifically, the probability of planting a particular crop one year depends on what was planted on the field in the prior year (Hendricks, et al., 2014a). We contribute to the previous and existing literature on forecasting crop production in two main ways: First, we generate crop choice transition probabilities based on first-order Markov theory using field-level data. Second, to the best of our knowledge, our acreage forecasting approach has not been employed and applied previously to forecast crop acreage distributions. Therefore, our novel approach will contribute to filling the gap in forecasting distribution of crop acreage.

2.4 Data

The primary source of information for crop choice data is the Cropland Data Layer (CDL). NASS produces the CDL, which was initiated in early 1997, to provide annual geospatial content to customers who were interested in annual cropland cover updates. CDL is a comprehensive, raster-formatted, and geo-referenced imagery for crop-specific land cover classification to identify field crop types accurately and geospatially (Boryan, et al., 2011).[9] The CDL includes the entire U.S. crop or land use classification codes, which are assigned to each pixel and classified by NASS using data from satellite sensors and validation. Use of the CDL data to date has been limited but has received more attention recently to study farmer's behavior regarding crop choice (Hendricks, et al., 2014a, Hendricks, et al., 2014b, Long, et al., 2014, Plourde, et al., 2013, Stern, et al., 2012, Yost, et al., 2014). For this manuscript, the Kentucky CLD data from 2010-2015 is the focus. Specifically, we examine the five main crops: corn, soybeans, tobacco, wheat, and alfalfa.[10] Kentucky provides an opportunity to examine a tradition corn-soybean rotation, a rotation for a contract crop in tobacco, and a perennial crop in alfalfa.

Next, we employ the Common Land Unit (CLU) boundaries, obtained from the GeoCommunity, to identify field boundaries.[11] Based on the Farm Service Agency (FSA) of the USDA, the CLU is defined as the smallest unit of land and individual contiguous farming parcel. The CLU is composed of contiguous boundary, common land cover, and land management (FSA, 2016).

To construct a filed level crop choice data, we used the following steps: First, we overlay the CLU with the National Land Cover Dataset (NLCD) 2011, which is the most recent national land cover product, produced by the Multi-Resolution Land Characteristics

---

[9] Raster, also called raster graphic, is simply an image that represents the rectangular grid of pixels. Each pixel in the CDL is a ground resolution of 30 meters by 30 meters.

[10] These are the top five row crops in Kentucky based on acreage. Please see the figure 1A in the appendix. This will provide an idea of how much areas these crops occupy in Kentucky.

[11] The CLU data was publicly available on FSA before 2008. However, FSA no longer provides the geospatial data including the CLU due to the Food, Conservation, and Energy Act of 2008. More detailed information about the CLU is available at https://www.fsa.usda.gov/programs-and-services/aerial-photography/imagery-products/common-land-unit-clu/index. In addition, more detail information about the GeoCommunity can be found at http://www.geocomm.com.

(MRLC) to remove non-agricultural fields.[12] Second, we overlay the CLU with the CDL to identify changes in rotations on a field by field basis instead of pixel or county basis. Third, we apply a moving window filter, which replaces each cell in raster based on the majority of adjacent cells, in Geographic Information System (ArcGIS) to remove misspecified (i.e., spurious) cells and to smooth rasters. Finally, we employ zonal statistics, which calculate the values of a raster within the zones of another dataset, to identify how many pixels are located in each field.[13] Table 2.1 shows the total observations and percent of observations by crop class and by year, respectively. We used 1,874,184 fields in total with approximately 1.5 million observations.[14] In 2015, the percentage of land in soybeans, corn, alfalfa, tobacco, and wheat acreage in Kentucky are 42%, 36%,1%, 0.6%, and 0.4%, respectively.[15] Twenty-eight (out of 120) counties in Kentucky are excluded from this study because the CLU data was not available. Table 2.2 shows the missing acres in percentage compared to the original CDL data. Based on Table 2.2, we are losing more data for tobacco and alfalfa compared to corn, soybeans, and wheat.[16]

Crop choice decisions by farmers are heavily dependent on the weather (e.g., precipitation and temperature) observed in the growing season. To control for weather factors, we obtained precipitation and temperature data from the Parameter-elevation Regressions on Independent Slopes Model (PRISM), which is official climatological data

[12] There are 16 different classifications of the NLCD, and this study only focuses on pasture/Hay and row crops as agricultural lands defined by the MRLC.
[13] More detail information about the zonal statistics can be found at http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-analyst-toolbox/h-how-zonal-statistics-works.htm
[14] This study encountered some difficulties in combining several public data sets with overlapping technical information on land characteristics and use. Appendix A describes what difficulties were encountered and how we overcame them.
[15] Figure 2.4 in Appendix C represents how these major crops are distributed in Kentucky. This figure shows that the majority of corn and soybeans are planted in the western Kentucky.
[16] Figure 2.5 in Appendix C shows the locations of excluded counties.

from the USDA.[17] For the weather variables, we calculate and use average precipitation in April, May, and June; the average temperature in June, July, and August months of each year.

Soil quality is another main factor that affects crop choice behavior by farmers. To control for this factor, we obtain data on soil textures (e.g., percent clay, percent silt, and percent sand) from Gridded Soil Survey Geographic (gSSURGO) database, which is provided by USDA National Resources Conservation Service (NRCS). The gSSURGO database has greater spatial extents (i.e., high resolutions) than the traditional SSURGO.[18] Finally, we obtain National Elevation data (30-meter resolution) from the Geospatial Data Gateway provided by USDA-NRCS to calculate the elevation and slope. The soil and elevation data sets are time-invariant.[19] All of the data, which are precipitation, temperature, slope, elevation, and soil textures, are spatially joined based on the unique field ID. Table 2.3 shows the summary statistics of the data used in this study. As shown in Table 2.3, the mean acreage of these fields is 5.97 acres, and their average soil composition is 66.25% silt, 20.59% clay, and 11.83% sand. These fields represent 92 counties in Kentucky (out of 120 counties).[20] In addition, the average monthly temperature and precipitation are 29.94 (°C) and 143.68 (mm), respectively.

[17] For more detail information about the PRISM dataset, see http://www.prism.oregonstate.edu/documents/PRISM_datasets.pdf
[18] For more detail information about the gSSURGO, see https://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs142p2_052164.pdf
[19] Soil and Elevation data sets in different years are not publicly available.
[20] 92 counties account for 32,149.43 square miles out of the total 40,407.78 square miles. Thus, our sample covers about 79.56% of total land in Kentucky.

2.5 Empirical Model

2.5.1 Multinomial Logit Model

This study employs the multinomial logit (MNL) model by McFadden (1973) to develop Markov transition probabilities for the five primary crops in Kentucky: corn, soybean, tobacco, wheat, and alfalfa. The MNL model is motivated by the random utility model (RUM) framework, and the following discussion of RUM is based on McFadden (1973) and Croissant (2012). The farmer chooses one alternative among different and exclusive alternatives. The decision to choose the alternative is then determined by the utility level, U_{ij}, for a farmer *i* derives from choosing alternative j. It follows that

$$U_{ij} = x'_{ij}\beta_j + \varepsilon_{ij} = V_{ij} + \varepsilon_{ij} \tag{1}$$

where $i = 1, \dots, N$, $j = 1, \dots, m$, V_{ij} is a deterministic component that depends on regressors and unknown parameters, ε_{ij} is an unobserved component (i.e., error terms).[21] This is called RUM, and the alternative providing the highest level of utility will be chosen; in other words, alternative j is chosen if and only if $\forall\ k \neq j\ U_j > U_k$. Suppose a farmer chooses at least one crop from different alternatives (other crops). The farmer will choose corn if the utility by choosing corn is higher than the utility by choosing soybeans (i.e., $U_{Corn} > U_{Soybean}$). Under the RUM framework, the utility and the choice are random in that some of the determinants of the utility are unobserved, implying the choice is supposed to be analyzed in probabilities. In this regard, we observe the outcome $y_i = j$ if alternative

[21] The regressors are case-specific regressors and alternative-specific regressor. The case-specific regressors vary over the farmer *i* but do not vary over the alternative *j* while the alternative-specific regressors vary over the farmer *i* and the alternative *j* (Green, 2003)

j provides the highest utility and the general expression of the probability of choosing alternative j can be defined as:

$$
\begin{aligned}
P_{ij} &= P(y_i = j) = P\big(U_{ij} > U_{ik} | x, \forall\, k \neq j\big) \\
&= P\big(\varepsilon_{ik} - \varepsilon_{ij} \leq x'_{ij}\beta_j - x'_{ij}\beta_k | x, \forall\, k \neq j\big)
\end{aligned}
\tag{2}
$$

According to Croissant (2012), the MNL model assumes that error terms are independently and identically distributed (IID). With this strong assumption, equation (2) can be shown that the choice probabilities are

$$
P_{ij} = P(y_i = j) = \frac{\exp(x'_{ij}\beta_j)}{\sum_{k=1}^{m} \exp(x'_{ik}\beta_k)} \tag{3}
$$

where P_{ij} , $0 < P_{ij} < 1$ and $\sum_{j=1}^{m} P_{ij} = 1$. In this study, we use crop choice at year t as the dependent variable and choice in year $t-1$, precipitation, temperature, soil texture variables, slope, and elevation as explanatory variables to model Kentucky farmers crop choice behavior. This study also considers and includes one more alternative, called "other." The choice of other is when the farmer does not plant any of five major crops in this field during our study period. The choice of other includes fallow, oats, barley, grain sorghum, and double-crop beans. In the MNL model, one set of coefficients is normalized to zero because there is more than one solution to set of coefficients (Greene, 2003). By setting $\beta_j = 0$, the set of coefficients corresponding to each outcome are estimated as following:[22]

[22] We record outcomes 1, 2, 3, 4, 5, and 6 for other, corn, soybean, tobacco, wheat, and alfalfa respectively. Since the recorded numerical values are arbitrary, greater number does not imply better outcome compared to the smaller number. In addition, outcome of no production (i.e., the choice of other) is our base outcome: $\beta_1 = 0$.

$$P(y_i = j) = \frac{\exp(x'_{ij}\beta_j)}{1 + \sum_{k=1}^{m} \exp(x'_{ik}\beta_k)} \tag{4}$$

2.5.2 Markov Chain Approach

The Markov chain approach has been widely used in land use studies such as Bell (1974), Baker (1989), Brown, et al. (2000), Muller and Middleton (1994), Savage (2011), and Thornton and Jones (1998). It has also been used in agricultural related studies especially for studying crop rotation behavior such as Aurbacher and Dabbert (2011), Castellazzi, et al. (2008), Matis, et al. (1985), Osman, et al. (2015), Paton, et al. (2014), and Troffaes and Paton (2013). Based on Taylor and Karlin (2014) and Savage (2011), a Markov process $\{X_t\}$ given the value of X_t is a stochastic process with the property that the values of X_u for $u < t$ do not affect the values of X_s for $s > t$. In other words, knowledge of past behavior is informative for the probability of any specific future behavior of the process if the current state is known. The Markov property in general can be defined as the following:

$$\Pr\{X_{t+1} = j | X_0 = i_0, \ldots, X_{t-1} = i_{t-1}, X_t = i\} = \Pr\{X_{t+1} = j | X_t = i\} \tag{5}$$

for all time periods *t* and all states $i_0, \ldots, i_{t-1}, i, j$.

The probability of X_{t+1} in state j given X_t in state i is called the one-step transition probability, denoted by $P_{ij}^{t,t+1}$. The Markov chain is stationary if the one-step transition probabilities are not a function of the time variable *t*. Considering stationary Markov chain, the one-step transition probabilities are re-written as

$$P_{ij}^{t,t+1} = P_{ij} = \Pr\{X_{t+1} = j | X_t = i\} \tag{6}$$

The one-step transition probabilities that outcome j observed in the current period given the outcome *i* observed in the previous period can be modified as follows:

$$P\{X_t = j | X_{t-1} = i_{t-1}, X_{t-2} = i_{t-2}, \dots\} = P\{X_t = j | X_{t-1} = i_{t-1}\} \quad (7)$$

With the stationary Markov chain, the transition probabilities are arrayed in a square matrix with a dimension based on the number of possible outcomes. The Markov process is then fully defined based on both the transition probability matrix and initial condition. In this study, we have six possible outcomes based on six different choices, so the transition probability matrix has 6×6 with elements, which are the probability of transitioning from the row outcome to the column outcome. A row outcome represents a crop choice among five crops in the previous year and a column outcome represents the crop choice in the current year.

2.6 Empirical Results

2.6.1 MNL Results and Transition Probabilities

Table 2.4 represents the estimated results of the MNL model. Based on the likelihood ratio chi-square of 443,752.37 with a p-value of 0.00 indicates that the model as a whole is statistically significant. Table 2.4 demonstrates the expected crop rotation results for all five crops used in this study. For example, in the corn case it is expected that there is a higher likelihood for soybeans or wheat to be planted after corn compared to corn, tobacco or alfalfa. We also calculate the marginal effects, which are reported in Table 2.5. Table 2.5 shows that if corn was planted, for example, in the previous year, corn and alfalfa are 13% and 0.2% less likely to be planed in the current year. However, soybean, tobacco, and wheat are more likely to be planted in the current year with 31.2%, 0.04%, 0.2%, respectively.[23] The marginal effects reflect traditional crop rotations as expected.

This study implements several hypotheses tests for the measures of fit. First, McFadden's R-squared shown in Table 2.4 from the MNL model, in general, does not provide a direct interpretation as does the R-squared in linear regression; however, it can be used to measure the goodness of fit for the MNL model. Based on Louviere, et al. (2000), the model fit is considered to be extremely good if the value of the McFadden's R-squared is between 0.2 and 0.4. Domenich and McFadden (1975) argue this range is equivalent to 0.7 to 0.9 for a linear regression model. Even though the McFadden's R-squared of 0.12 presented in Table 2.4 does not provide strong evidence of extremely good model fit, it does not mean that our model is inappropriate. Second, we test whether all of the coefficients associated with the independent variable are equal to zero by using the Likelihood-ratio (LR) test and Wald test. These tests allow us to determine whether the independent variables used in the MNL model are significant across all outcome categories. Based on results of both tests, we reject the null hypothesis that all coefficients associated with given variables are zero, implying no variables can be dropped from the model since independent variables have a significant impact across all crop choices. Third, we test whether some categories of the dependent variable can be combined or not by using the Wald test. If outcomes are not differentiated concerning the independent variables, we combine outcomes. We find that any pair of outcome categories cannot be combined by

[23] In table 2.5, we do not report the marginal effect for choice of other. In general, the marginal effects sum up to zero.

rejecting the null hypotheses that independent variables do not differentiate between outcome categories.

Furthermore, this study conducts a validation test for the predicted probabilities from the MNL model. This will measure accuracy of predictions on crop choice using the estimated MNL model. For the validation test, we employ the following steps. First, we select 10% of the data using a random sampling process, called a test sample. Second, we estimate the MNL model with the remaining dataset (90% of data, also called training set), and then predict using the test sample. Third, we assign crops based on the highest predicted probabilities for each field. For example, a field X using MNL model is predicted with 55%, 20%, 10%, 5%, 5%, and 5% probabilities for corn, soybeans, tobacco, alfalfa, wheat, and others, respectively. The corn is then assigned to the field X because corn has the highest probability of being planted. Finally, we compare the predicted and actual crop choice in each field and calculate the accuracy percentage, which is a number of accurate predictions over the total number of predictions. From the out of sample validation method, this study finds the probabilities are predicted with 51.9% accuracy. We also resize the test sample sizes by 20% and 30%, and we find that probabilities are correctly predicted by 52.2% and 52.2%, respectively. Even though the probabilities from the MNL model are predicted with approximately 52% accuracy, the predicted probabilities for the crop choice transition matrix and forecast distribution are plausible.[24]

Based on the results from the MNL model in Table 2.4, we generate a set of average conditional predicted probabilities or Markov transition probabilities. Table 2.6 shows

[24] From the discrete choice model standpoint, the probabilities will be better predicted with fewer outcome categories. Since we have six different outcome categories, out of sample prediction with 52% accuracy is considered as a reliable prediction.

these calculated as the average number of observed transitions between 2010 and 2015.[25] Based on Table 2.6, this study finds that if corn is planted in year *t*, there is a 22% chance that corn will be planted in year *t*+1. Crop rotation probabilities between corn to soybean and soybean to corn from the current year to the next year are 46.2% and 58.7%, respectively. Compared to other crops, such as tobacco, wheat, and alfalfa, transition probabilities of corn and soybean show relatively lower probabilities in their transition. Martinez and Maier (2014) state that crop rotation between cereal crops such as corn and wheat followed by leguminous crops such as soybean and alfalfa is a common example. Therefore, farmers switch corn to soybean and soybean to corn for not only maintaining and improving the soil fertility but also protecting the environment from the nitrogen runoff.

2.6.2 Simulation Exercise

The objective of this section is to generate a distribution by reflecting total acreage of crop *i* for the year *t* using information up to year *t*-1 (e.g., predicting crop composition in 2016 using data up to 2015). For this purpose, we follow three steps. First, we used a multinomial logit model specified in the previous section to estimate the probability of field *j* with crop *i*. Using this method, we generate a matrix of probabilities that have a probability of crop *i* (where *i* includes corn, soybeans, tobacco, wheat, alfalfa, and other crops) being planted in field *j* where the sum of each row would be 1. Second, we utilize random sampling with 1,000 iterations to identify crop choice in each field based on the transition probability matrix. In each iteration, field *j* is assigned to crop *i*. The average

[25] Table 2.7 in the Appendix B provides the yearly transition probabilities in percentage.

probability of choosing crop *i* for field *j* after 1,000 iterations would be approximately the same as the choice probabilities that are used in the first step. Up to this point, we forecasted the crop choice by each field in year *t*. Third, we assume that acreage that is going to be planted in year *t* is the same as the planted acreage at year *t*-1. Based on this assumption, we assigned planted acreage in each field at year *t*-1 to the forecasted crop. This process generates forecasted acreage for each crop and each field for year *t*.

This process is replicated for each iteration (1,000 times). Next, for each iteration and each crop, we calculate aggregate acreage that would be planted in year *t*. Finally, we used the forecasted aggregate acreages for each iteration and crop to calculate the distribution of total expected acres in year *t*.

We apply this method to Kentucky fields and predict the expected crop composition in 2016 using realized crop choice from 2010 to 2015. Figure 2.1 indicates forecasted distributions for each crop in 2016. When comparing these to the actual state average acres produced, we find that the forecasted mean is close to the historical means for the simulated counties. There are missing counties from the data set due to the lack of CLUs for those counties. One word of caution when interpreting this is to watch the scales on the x-axis. The two largest crops produced in the state are by far corn and soybeans. As expected, these distributions are significantly wider than the other three crops considered. In general, there is a little variation in the tobacco, wheat and alfalfa acres in the simulation. There are several reasons for this result. First, alfalfa is a perennial crop and has a five to seven-year rotation, and is typically rotated with tobacco. This causes the acres of this crop to be stable over time. Secondly, tobacco is primarily produced via a production contract. This creates a situation where the acreage will be quite stable from year to year. Thirdly, for wheat, Kentucky is home to significant milling and distilling industries that contract with producers to ensure supply. Lastly, this leaves corn and soybean acres to be the primary acreages in the state that shift to fill in the gaps. These distributions demonstrate this phenomenon.

In addition to the separate crop distribution in figure 2.1, we generate a violin distribution for each crop to compare shapes of the distributions between crops and the distributions are presented in figure 2.2. These distributions are a different way of presenting figure 2.1 in which we can compare differences in forecasted acres between the crops. In this figure, the top panel compares the forecasted crop acreage distribution between corn and soybeans, and bottom panel compares forecasted acres between tobacco, wheat, and alfalfa. Using this figure, we can identify that the forecasted acreage of corn is the highest among other crops followed by soybeans, and tobacco and wheat are the lowest, implying corn and soybean plantings dominate in Kentucky. In addition, the widths of the distributions provide insights into where crops such as hemp would enter the crop rotation. The wider the distribution, the more likely acreages from these crops are to shift.

Furthermore, we generate a distribution of each crop sales based on the forecasted distributions, which are presented in figure 2.3. For example, sales of corn, soybeans, alfalfa, wheat, and tobacco, on average, are \$599, \$400, \$8.7, \$1.24, and \$18.04 (in million dollars), respectively in 2016.

To review the main results, there are higher probabilities of planting soybeans or wheat after corn relative to corn after corn, tobacco, or alfalfa. In addition, the transition probability of the crop rotation demonstrates that corn will be planted after soybean, and vice versa, and that alfalfa has a lower probability of being rotated with other crops from year to year. These findings are expected with traditional crop rotation in the U.S., and a characteristic of a perennial crop, especially for alfalfa. Finally, forecasting results indicate that there are significantly wider distributions in corn and soybean, whereas there is little variation in the tobacco, wheat and alfalfa acres in the simulation.

2.7 Concluding Remarks and Policy Implications

This study proposes a novel application of the multinomial logit (MNL) model to estimate the conditional transition probabilities of crop choice and forecast distributions of total acreages by crop type using recovered transition probabilities. For this purpose, we utilize the Cropland Data Layer (CDL), which is overlayed with the Common Land Unit (CLU) dataset to identify crop choice at the field-level accurately. In this paper, we focused on the production of corn, soybeans, tobacco, wheat, and alfalfa in Kentucky from 2010 to 2015.

Based on transition probability estimation results, we find that corn is more likely to be followed by soybeans, as would be expected. For tobacco and alfalfa, they are found to be monoculture crops since they are more likely to be planted in consecutive years. These findings are consistent with the traditional crop rotation in Kentucky. Our forecasted distributions based on the simulation exercise show wider distributions for corn and soybeans, whereas narrower for tobacco, wheat, and alfalfa. The different shapes of the distribution can be explained in that alfalfa is a perennial crop and tobacco is a contracted crop. The forecasted distributions can be used and applied in various fields of research and will contribute to policy implications.

Kentucky, for example, is the largest cow/calf state east of the Mississippi River and one of the reasons for this is its ability to produce forages or biomass. In the period between 2010 and 2014, a large number of acres in the state switched from pastureland to crop production to take advantage of record prices. Now that crop prices are trending down, many of these producers are searching for alternative viable feedstocks. Therefore, the identification of viable feedstocks and locations that these feedstocks can be produced is one of the important steps in the development of the Kentucky bioeconomy.

The method proposed in this study could be used to evaluate where the most likely places are for the production of industrial hemp, but it could also be used to evaluate other potential crops, as they all have to compete for scarce land resources. By recovering the forecast distributions of the traditional commodities grown, such as corn, soybeans, wheat, tobacco, and alfalfa, producers have a much richer view of the future when they decide to adopt alternative feedstocks. Moreover, the forecasted crop acreage distribution can support management decisions for fertilization, irrigation, and pesticide uses. The crop acreage distribution could further provide the basis for planning, formulation, and implementation of policies related to the crop procurement, distribution, price structure, and import-export decisions. Our results may also supplement NASS survey in areas where response rates are low and could serve as estimates in the winter season before the Spring Survey. The generation of the forecast distributions could be one way for farmers, policy-makers, and other stakeholders to consider uncertainty in forecast estimates by crop.[26]

[26] The historical distributions of Corn and Soybeans are presented in Figure 2.6 in Appendix C. We find that the forecasted distribution of Corn looks similar to the historical distribution of Corn even though the average acre is a little lower in the forecasted distribution. This can be explained by missing counties in CLU (see figure 2.5 in Appendix C). For soybeans, we find that the historical distribution of soybeans is right skewed, whereas the forecasted distribution of soybean is normal. This indicates that the average acreage reported in NASS report tends to be overly provided.

Finally, further expansion would be the identification of critical thresholds based on the forecasted distributions that can be applied not only to calculate the total nitrogen or fertilizer runoff for sustainable agriculture but also to develop wildlife habitat management plan.

Several limitations should be outlined alongside our findings. First, our state-level aggregated prediction results might be under- or over-estimates due to the 28 missing counties and accuracy of CDL data. Some crops are poorly identified in the CDL data. Based on CDL accuracy assessment information provided by NASS, average accuracy for tobacco, wheat, and alfalfa in Kentucky are 76.1%, 48.8%, and 74.1% from 2010 to 2015, whereas corn and soybeans are identified with 96.1% and 93.5% accuracy, respectively. Higher quality CDL data will result in better predictions. In this study, we consider corn and soybeans as our best estimates since they have the highest accuracy of the CDL data. Second, in this study, for simplicity and data availability, we assume that farmers plant a single crop in each field per growing season instead of planting several crops in one field in one growing season. In reality, however, a farmer might grow more than one crop in their field which needs to be addressed in the future studies. Third, our MNL model and simulation exercise are only based on agronomy-based characteristics. Therefore, this study can be extended by incorporating microeconomic variables such as expected net return, expected price, and farmer's characteristics. Finally, this study focuses on Kentucky, where agricultural lands are relatively heterogeneous compared to some states like Iowa, Illinois, and Nebraska where agricultural lands are homogeneous. Our results may not be consistent with those states with the homogeneous agricultural lands; therefore, future research will be needed by looking at other states.

2.8 Tables and Figures

Table 2.1 Total Number of Fields in the Sample by Crop Class by Year

	2010	2011	2012	2013	2014	2015	Sum
Corn	105,266 (33.70)	100,060 (32.03)	123,638 (39.58)	112,930 (36.15)	113,079 (36.20)	113,662 (36.39)	668,635
Soybean	98,311 (31.47)	79,480 (25.44)	76,268 (24.41)	79,492 (25.45)	99,343 (31.80)	129,684 (41.51)	562,578
Tobacco	456 (0.15)	298 (0.10)	734 (0.23)	1,056 (0.34)	1,317 (0.42)	1,730 (0.55)	5,591
Wheat	276 (0.09)	254 (0.08	267 (0.09)	344 (0.11)	203 (0.06)	1,232 (0.39)	2,576
Alfalfa	1,742 (0.56)	2,936 (0.94	1,593 (0.51)	1,537 (0.49)	2,168 (0.69)	3,136 (1.00)	13,112
Other	106,313 (34.03)	129,336 (41.40)	109,864 (35.16)	117,005 (37.45)	96,254 (30.81)	62,920 (20.14)	621,692
Sum	312,364	312,364	312,364	312,364	312,364	312,364	1,874,184

Notes: Numbers in parenthesis are the percent of fields by crop class by year. These numbers are calculated as dividing a number of fields for each crop by the total number of fields in each year (last row in table 2.1).

Table 2.2 Percentage of Missing Acres in Out Analysis Based on Missing Counties in the CLU by Crop Type and Year

Year	Corn	Soybean	Wheat	Tobacco	Alfalfa
2010	19.45%	17.73%	52.43%	64.44%	68.29%
2011	19.66%	20.16%	32.60%	44.99%	55.79%
2012	19.37%	20.86%	29.49%	39.69%	58.53%
2013	19.58%	19.84%	47.87%	33.40%	56.37%
2014	19.67%	21.81%	44.47%	33.89%	62.02%
2015	20.44%	21.74%	30.31%	31.41%	57.58%

Notes: The field boundary in CLU has 28 missing counties out of 120 counties in Kentucky. We compared the observed acres based on CLU and the acres from the CDL. For example, 19.45% of corn acres in 2010 are missing in our analysis.

Table 2.3 Summary of Data Used for Estimation of Conditional Probabilities

Variable	Mean	S.D.	Min.	Max.
Acres	5.97	15.92	0.00	934.28
Silt (%)	66.25	13.27	0.00	82.00
Clay (%)	20.59	7.00	0.00	58.00
Sand (%)	11.83	8.75	0.00	93.90
Slope	2.51	1.95	0.00	35.07
Elevation (meter)	174.79	60.56	0.00	423.09
Temperature (Celsius)	24.94	1.27	20.78	28.13
Precipitation (mm)	143.68	55.25	0.00	308.89

Notes: Summary statistics are based on the observations from 2010 to 2015. Units of the observations are reported in the parenthesis. Acres indicate a number of acreages planted, and zero acreage refers to another choice. S.D represents standard deviation.

Table 2.4 Conditional Multinomial Logit Model Results

	Corn	Soybean	Tobacco	Wheat	Alfalfa
Choice lag					
Corn	-0.013***	1.580***	0.294***	1.806***	-1.065***
	(0.005)	(0.005)	(0.035)	(0.055)	(0.039)
Soybean	1.916***	2.041***	1.016***	1.866***	-0.220***
	(0.006)	(0.007)	(0.040)	(0.067)	(0.040)
Tobacco	0.210***	0.984***	4.501***	2.517***	0.112
	(0.044)	(0.044)	(0.053)	(0.212)	(0.210)
Wheat	1.176***	0.927***	1.625***	3.001***	2.038***
	(0.064)	(0.085)	(0.309)	(0.311)	(0.157)
Alfalfa	-0.975***	-0.618***	-0.409**	0.418	3.691***
	(0.038)	(0.045)	(0.220)	(0.294)	(0.031)
Observations:	1,561,820				
McFadden R-Square:	0.123				
Log Likelihood:	-1,587,530				

Notes: Robust standard errors are reported in parenthesis. Control variables used in this model are crop acreage, soil type (slit, clay, and sand), slope and elevation at the filed-level, average monthly temperature, and total monthly precipitation. Significance levels are indicated by ***, **, * for 10, 5, and 1 percent significance level, respectively.

Table 2.5 Marginal Effect from Conditional Multinomial Logit Model Results

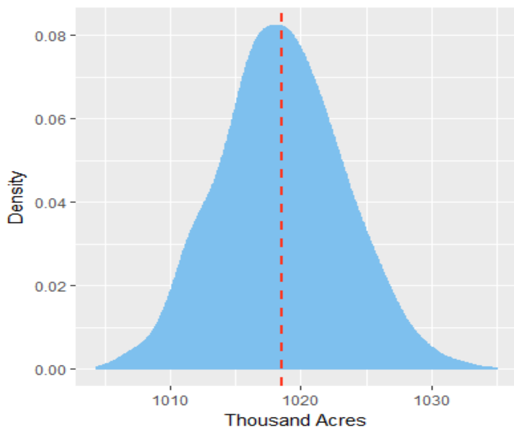
	Corn	Soybean	Tobacco	Wheat	Alfalfa
Choice lag					
Corn	-0.130***	0.312***	-0.0004***	0.002***	-0.002***
	(0.001)	(0.001)	(0.0001)	(0.0001)	(0.0001)
Soybean	0.239***	0.137***	-0.0007***	0.0004***	-0.002***
	(0.001)	(0.001)	(0.0001)	(0.0001)	(0.0001)
Tobacco	-0.071***	0.109***	0.133***	0.004***	-0.001*
	(0.008)	(0.007)	(0.005)	(0.001)	(0.0004)
Wheat	0.203***	0.035***	0.003**	0.005***	0.008***
	(0.014)	(0.011)	(0.002)	(0.002)	(0.002)
Alfalfa	-0.188***	-0.052***	-0.0005	0.001	0.132***
	(0.005)	(0.004)	(0.0004)	(0.0003)	(0.003)

Notes: Robust standard errors are reported in parenthesis. Marginal effect for the control variables used in this model are not reported here. Significance levels are indicated by ***, **, * for 10, 5, and 1 percent significance level, respectively.

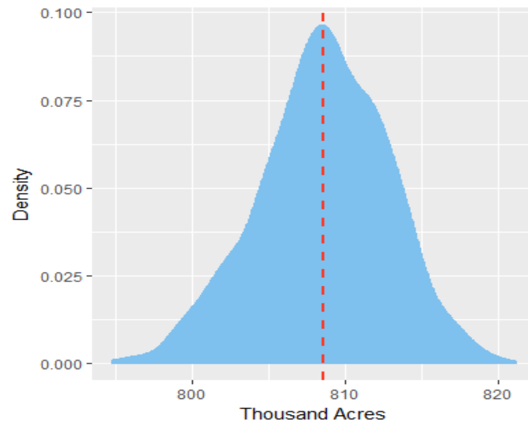
Table 2.6 Conditional Transition Probabilities (in percentage)

		To:					
		Other	Corn	Soybean	Tobacco	Wheat	Alfalfa
From:	Other	51.07	32.69	14.72	0.38	0.08	1.07
	Corn	30.80	22.40	46.17	0.25	0.24	0.14
	Soybean	12.20	58.67	28.66	0.20	0.11	0.16
	Tobacco	32.12	23.59	23.75	19.32	0.62	0.60
	Wheat	28.42	49.11	17.04	0.82	0.82	3.79
	Alfalfa	46.61	8.56	5.85	0.21	0.12	38.64

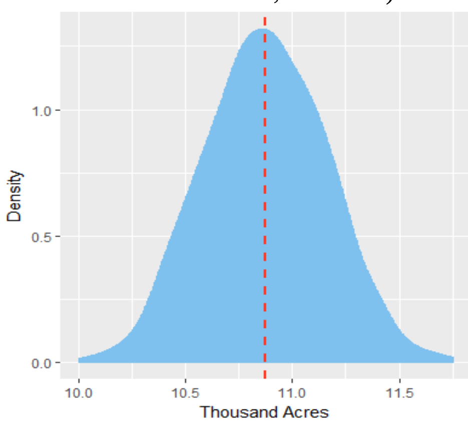
Notes: Estimated conditional transition probabilities are based on the conditional multinomial logit model. Transition probabilities are presented in percentage form (e.g., transition probability from corn to corn is 22.40%). In this table, "Others" include fallow, oats, barley, grain sorghum, and double-crop beans.



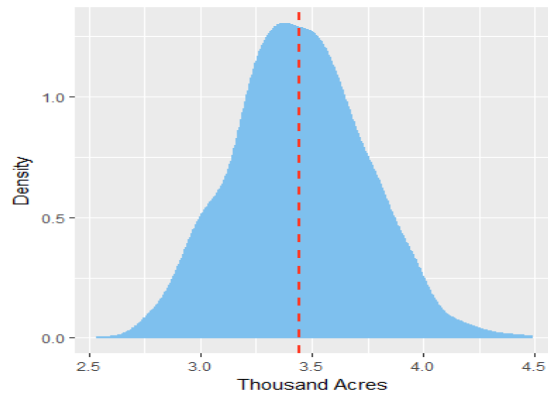
Corn (mean=1,019; Min.=1,004; Max.= 1,035, SD= 4.75, CV=0.47)



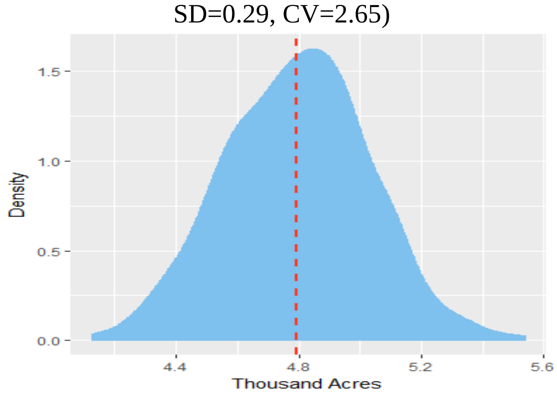
Soybeans (mean=809; Min.=795; Max.= 821, SD= 4.26, CV=0.53)



Alfalfa (mean=1.87; Min.=10; Max.= 11.75, SD=0.29, CV=2.65)



Wheat (mean=3.44; Min.=2.53; Max.= 4.48, SD= 0.29, CV=8.56,)



Tobacco (mean=4.79; Min.=4.12; Max.= 5.53, SD=0.24, CV= 4.96)

Figure 2.1 Forecasted distributions for each crop in the year 2016. Dashed red lines indicate mean forecasted acreage in 2016. Summary statistics of the distributions are reported in parenthesis, and all units are in thousands of acres.

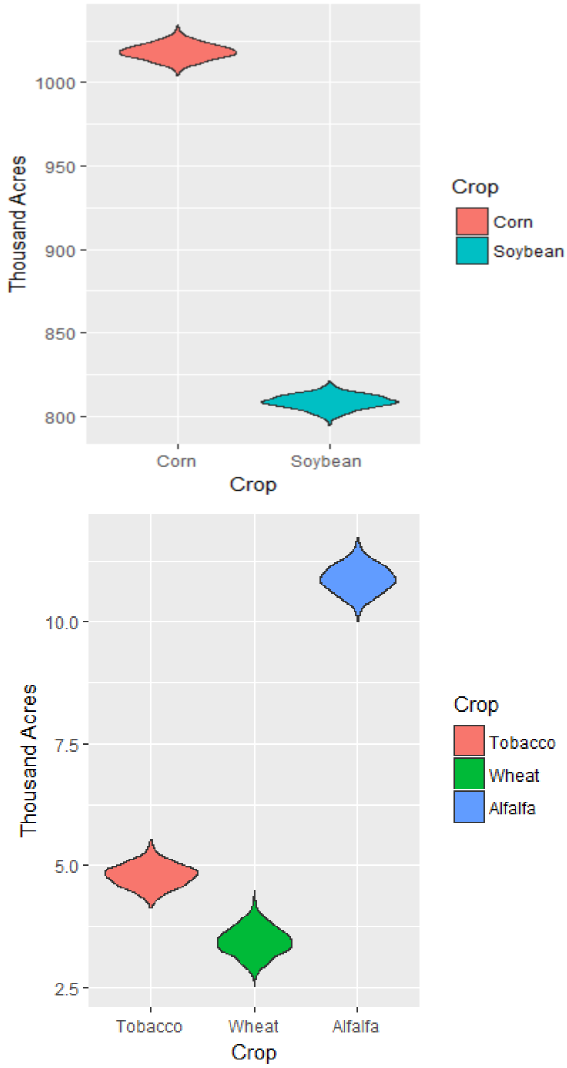
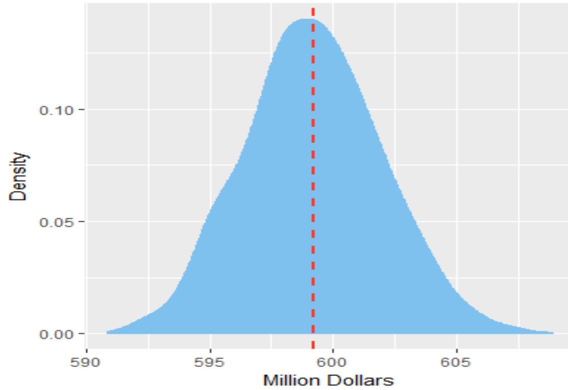
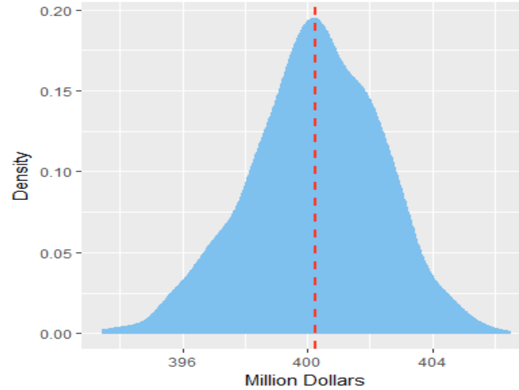


Figure 2.2 Comparison of forecasted distributions of crop acreage in 2016 in Kentucky.

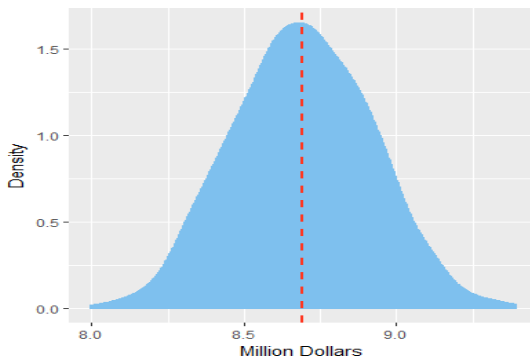
Notes: Major crops divided into two groups: group one is Corn, Soybeans, and group two includes Tobacco, Wheat, and Alfalfa.



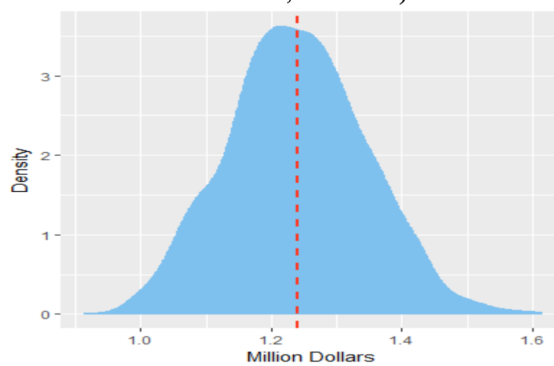
Corn (mean=599; Min.=591; Max.= 609, SD=2.79, CV=0.47)



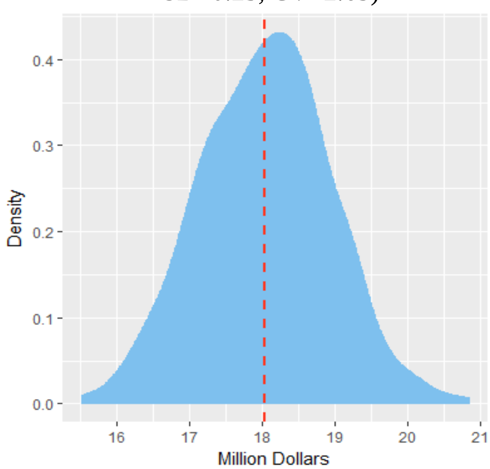
Soybeans (mean=$400; Min.=393; Max.= 406, CD=2.11, CV=0.53)



Alfalfa (mean=8.7; Min.=7.99; Max.= 9.39, SD=0.23, CV=2.65)



Wheat (mean=1.24; Min.=0.91; Max.= 1.61, SD=0.11, CV=8.56)



Tobacco (mean=18.04; Min.=15.53; Max.= 20.84, SD=0.89, CV= 4.96)

Figure 2.3 Forecasted distributions for each crop sales (million dollars) in 2016.

Notes: Dashed red lines indicate mean forecasted acreage in 2016. Summary statistics of the distributions are reported in parenthesis, and all units are in a million dollars.

2.9 Appendix A: Discussion of Data Set

Although we mentioned and described how we merged and utilized all the different data sets in the data section, we would like to discuss some difficulties we had and how we overcame them more in detail. First, we realized that the field boundary data do not cover the entire state of Kentucky (i.e., there are missing field boundaries) as we mentioned in the manuscript (see Figure 2B). However, we think those missing fields do not represent the major crops produced in Kentucky. In other words, those missing fields may be a problem in other states if the missing fields cover a large portion of the major croplands. Utilizing the CDL data with CLU, therefore, researchers should pay attention to those missing fields and carefully check before merging with the CDL. Second, we had an issue in identifying the crop choice. In reality, each field is not fully covered by one single crop based on a pixel in CDL when we overlay with CLU (see Figure 1B). For instance, there might exist multiple crops in one single field, so we make a strong assumption that the field is corn if corn acreage dominates other crops. Alternatively, the analyst is able to identify the representative crop in the field by using a centroid point if the point interacts spatially with the pixel. This alternative method might work with a large sample such as all 48 contiguous U.S. states instead of one single state. However, this may provide inaccurate results from a forecasting perspective. Third, some of the data sets (especially Soil, Slope, and Elevation data) used in this study are not time-varying data. However, those data sets are only available information that can be merged with the field level data, and we assume soil quality, slope, and elevation for lands do not significantly vary over time.

2.10 Appendix B: Additional Table

Table 2.7 Year to Year Conditional Transition Probabilities (in percentage)

Year	Choice	Other	Corn	Soybean	Tobacco	Wheat	Alfalfa
2011	Other	48.75	29.01	21.01	0.23	0.12	0.88
	Corn	29.64	44.05	25.58	0.18	0.11	0.43
	Soybean	32.85	33.88	32.55	0.18	0.15	0.39
	Tobacco	44.15	30.82	21.67	2.16	0.19	1.01
	Wheat	38.07	29.41	31.01	0.39	0.21	0.90
	Alfalfa	53.74	20.30	12.42	0.38	0.13	13.02
2012	Other	43.84	34.45	20.73	0.34	0.11	0.53
	Corn	27.89	48.86	22.67	0.29	0.09	0.19
	Soybean	28.78	39.41	31.17	0.30	0.14	0.19
	Tobacco	37.63	37.66	22.08	2.05	0.16	0.42
	Wheat	36.35	33.15	29.60	0.38	0.18	0.34
	Alfalfa	49.79	22.11	12.69	0.53	0.15	14.72
2013	Other	38.95	28.47	30.96	0.38	0.17	1.06
	Corn	26.05	39.57	33.27	0.34	0.15	0.61
	Soybean	24.66	30.21	44.06	0.34	0.22	0.51
	Tobacco	30.64	30.61	34.98	2.48	0.26	1.04
	Wheat	30.04	28.20	39.80	0.70	0.28	0.98
	Alfalfa	39.90	18.80	15.63	0.54	0.17	24.95
2014	Other	41.33	29.81	27.50	0.38	0.15	0.83
	Corn	27.58	41.59	29.90	0.34	0.13	0.45
	Soybean	26.67	32.54	39.71	0.39	0.20	0.49
	Tobacco	33.88	31.33	29.52	3.95	0.26	1.06
	Wheat	34.24	28.41	34.57	0.61	0.31	1.86
	Alfalfa	41.91	21.51	15.99	0.60	0.17	19.81
2015	Other	38.73	30.61	29.23	0.33	0.15	0.95
	Corn	27.68	42.89	28.40	0.32	0.13	0.57
	Soybean	28.06	33.78	36.92	0.36	0.19	0.68
	Tobacco	38.20	29.65	25.35	5.05	0.28	1.47
	Wheat	29.49	29.69	39.08	0.68	0.26	0.81
	Alfalfa	39.21	20.12	16.03	0.60	0.18	23.85

Notes: Transition probabilities are presented in percentage form. For example, 44.05% of corn will be planted in 2011 if the corn was planted in 2010.

2.11 Appendix C: Additional Figures

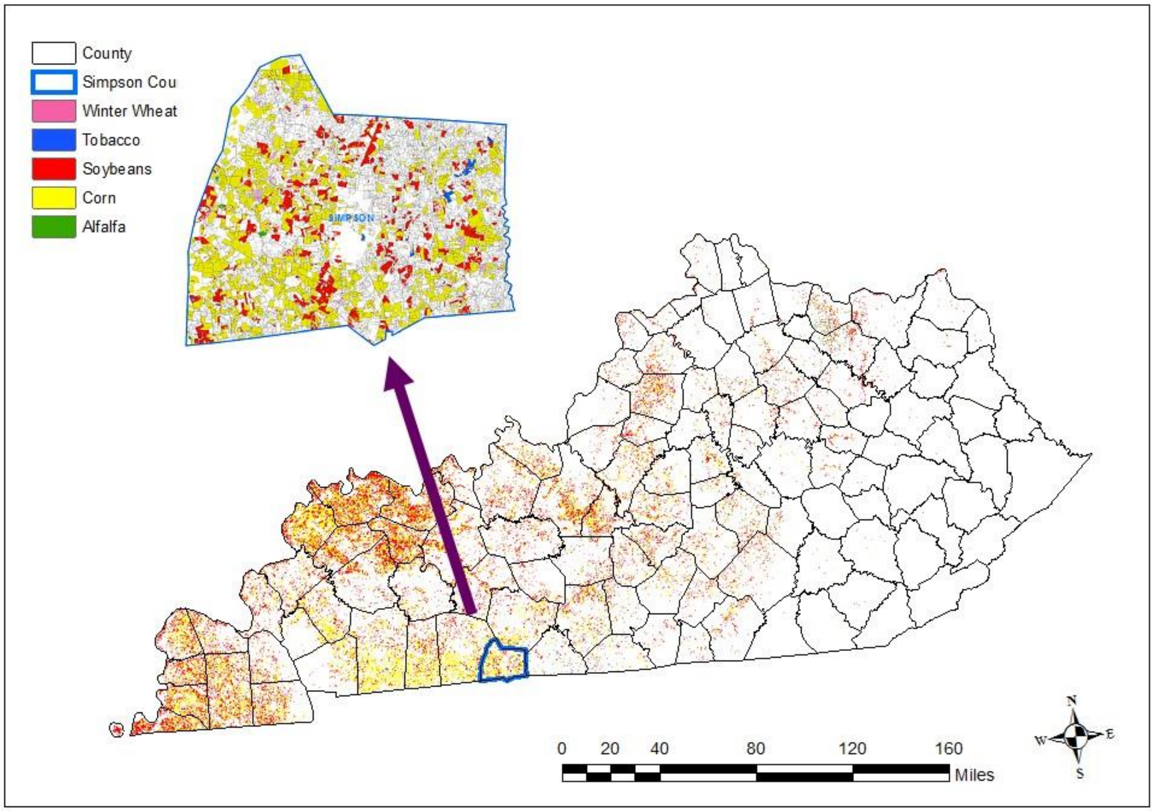
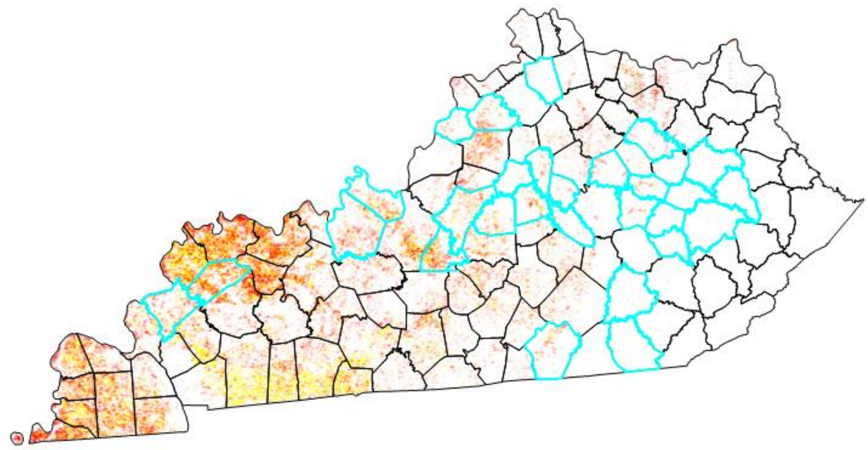
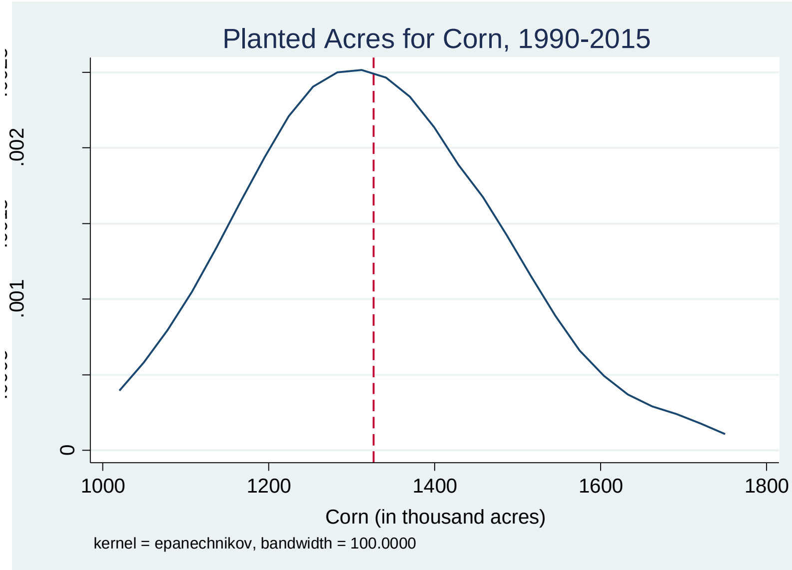


Figure 2.4 2015 major crops produced in Kentucky from CDL

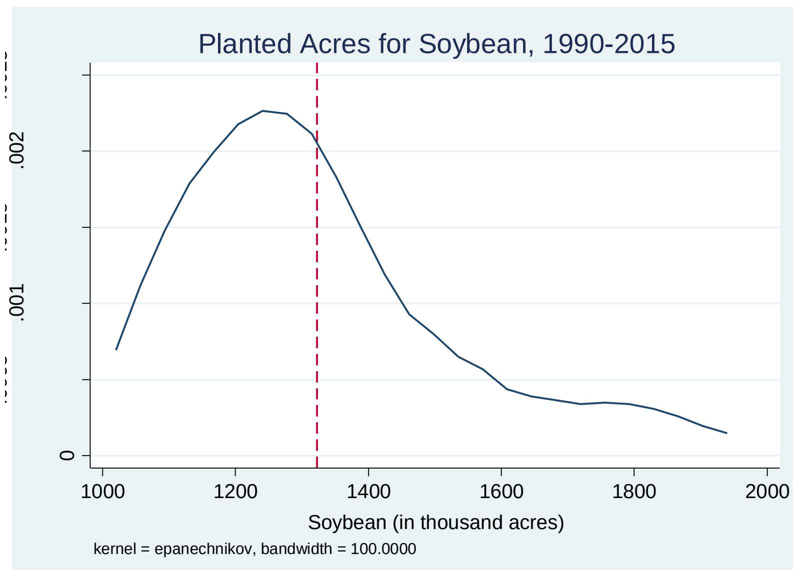


County	FIPS
ANDERSON	5
BATH	11
BREATHITT	25
BRECKINRIDGE	27
CLARK	49
CRITTENDEN	55
ESTILL	65
GARRARD	79
GRANT	81
HENRY	103
JESSAMINE	113
LARUE	123
LAUREL	125
LEE	129
MAGOFFIN	153
MEADE	163
MERCER	167
MORGAN	175
NELSON	179
OLDHAM	185
OWSLEY	189
POWELL	197
WASHINGTON	229
WAYNE	231
WEBSTER	233
WHITLEY	235
WOLFE	237
WOODFORD	239

Figure 2.5 Numbers of Missing Counties in Kentucky based on CLU



Corn (Mean = 1,326.5 Std. Dev = 123.9 Min = 1,120 Max = 1,650)



Soybeans (Mean = 1,322.7 Std. Dev = 192.9 Min = 1,120 Max = 1,840)

Figure 2.6 Historical Distributions of Corn and Soybeans from 1990 to 2015

CHAPTER 3. PROFILING CONSUMER OF HEMP FOODS IN THE U.S.: EVIDENCE FROM NIELSEN CONSUMER PANEL DATASET FROM 2008 TO 2015

3.1 Abstract

The objective of this study is to identify critical consumer-demographic characteristics that are associated with the consumption of products containing hemp and investigate their effect on total expenditure in the U.S. To estimate the likelihood of market participation and consumption level, the Heckman selection model, is employed using the maximum likelihood estimation procedure utilizing Nielsen consumer panel data from 2008 to 2015. Results indicate marketing strategies targeting consumers with higher education and income levels can attract new customers and increase sales from current consumers for this burgeoning market. Head-of-household age in different regions shows mixed effects on decisions to purchase hemp products and consumption levels. Findings will provide a basic understanding of a consumer profile and overall hemp market that has had double-digit growth over the last six years. As the industry continues to move forward, policymakers are going to need a deeper understanding of the factors driving the industry if they are going to create regulations that support the development of the industry.

3.2 Introduction

Can a market that is expected to top $1.8 billion in sales by 2020 be based on a feedstock that was classified as a schedule 1 narcotic less than a few month ago? 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. Industrial hemp is a variety of the *Cannabis sativa* plant species with delta-9 tetrahydrocannabinol concentration (THC) of no more than 0.3 percent on a dry weight basis.[27] 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). The Comprehensive Drug Abuse

[27] See more detail about the 2014 Farm bill at https://www.gpo.gov/fdsys/pkg/BILLS-113hr2642enr/pdf/BILLS-113hr2642enr.pdf

Prevention and Control Act in 1970 classified industrial hemp as a schedule 1 narcotic. In this regards, growing industrial hemp in the U.S. for commercial purpose was illegal and strictly restricted.[28] Within the U.S. Section 7606 of the 2014 Farm Bill set the reintroduction of industrial hemp as a potential crop to be grown in motion. Interest in this crop has continued to gain momentum with acreage growing to approximately 40,000 acres in 2017. On December 2018, the 2018 Farm Bill was approved by U.S. Congress, and it allows industrial hemp to be legalized by removing it from the Controlled Substances Act.[29] To be specific, industrial hemp is allowed not only to cultivate broadly but also to produce hemp-derived products across state lines based on the 2018 Farm Bill.

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 are produced around the world each year. Approximately 70 percent of industrial hemp in the world is produced in China, Russian, and South Korea.[30] According to Fortenbery and Bennett (2004), industrial hemp production has 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 industrial hemp on the demand side are increased efficiency compared to other inputs for industrial use, health benefits of both hemp oil and hemp

---

[28] Although growing industrial hemp was illegal in the U.S., some states such as Colorado, Kentucky, and North Carolina grew and produced as a research or pilot programs.
[29] https://www.fda.gov/ohrms/dockets/ac/03/briefing/3978B1_07_A-FDA-Tab%206.pdf
[30] http://www.kyagr.com/marketing/industrial-hemp.html

hearts consumption, and competitive use in textile manufacturing (Fortenbery and Mick, 2014).[31]

Since there is minimal commercial production in the U.S. due to production restrictions, most hemp-based products are imported from other countries.[32] For instance, raw and processed hemp fiber is predominantly imported from China whereas hemp seed and oilcake are imported from Canada (Johnson, 2017). Figure 3.1 provides the total value of U.S. hemp imports from 2010 to 2015, and it shows that the total value of imported hemp is increasing. Based on Johnson (2017), Hemp Industries Association (HIA) estimates that annual growth in U.S. hemp retail sales averaged more than 15% from 2010 to 2015. The author also mentions that 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 at approximately $688 million including food and body products, clothing, auto parts, building materials, and other products.[33]

[31] For example, hemp can be substitutable for cotton to make textile in that hemp fiber is 10 times stronger than cotton; in addition, hemp can be used as building materials instead of wood at low manufacturing costs. More detail environmental and economic benefits of hemp can be found at http://www.nemeton.com/static/nemeton/axis-mutatis/hemp.html

[32] Even though the process of commercial production in the U.S. has been started as a pilot program for the research purpose since 2014, growing hemp is still illegal in the U.S. according to federal law.

[33] 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

Even though there is minimal commercial hemp production in the U.S., retail sales for hemp production is increasing over time. Based on our best knowledge, no study has investigated and examined factors that affect consumption of hemp products. In this study, we investigate the critical economic and demographic characteristics that are associated with hemp consumption and investigate their effects on expenditures in the U.S. by utilizing Nielsen's consumer panel data from 2008 to 2015. Due to the limited data availability, we limit the investigation to four different categories of hemp containing products including granola, nuts, nutrition, and protein.[34] This study employs a 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 total expenditures on hemp products. Furthermore, we take account of states that have passed regulations that allow the production of industrial hemp. The hypothesis is that the probability of purchasing and expenditures on hemp products are relatively higher in states that have already passed this legislation.[35]

Findings from this study will contribute to understanding the continued growth of the burgeoning industrial hemp market as the U.S. Congress approved 2018 Farm Bill to allow the commercialization of hemp production. Also, this paper provides potential market strategies by not only understanding consumers but also targeting groups of

[34] The term of hemp product used in this study is referred to the product that contains hemp. However, we make no designation as to the amount of hemp contained in the products. It could be .0001% to 100%. Later in the paper, we refer to the four different hemp products as hemp granola, hemp nuts, hemp nutrition, and protein in order to avoid confusion.
[35] In this study, hemp legislation infers any legislations that have been passed or introduced in the state to allow commercial hemp.

consumers to increase the market share of 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 econometric model. The following section describes the data section especially the structure of the data and the variable classification. The next section presents results and discussions, while the final section summarizes the main results with limitations and directions for future research.

3.3 Background

3.3.1 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 an essential crop in Colonial America since New England first grew the plants for a 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 for use in 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, hemp production began to decline 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 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 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), hemp production reached more than 150 million pounds on 146,200 harvested acres in 1943, 140.7 million pounds were hemp fiber, and 10.7 million pounds were hemp seed. However, hemp production declined to 3 million pounds on 2,800 harvested acres in 1948. The decline in 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, there has been virtually no U.S. hemp production since then (Dempsey, 1975, Ehrensing, 1998, Fortenbery and Mick, 2014).

3.3.2 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 federal laws and drug policy have restricted domestic hemp production 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). In the mid-1990s, hemp resurfaced in the U.S. as the potential uses of the plant expanded and after Europe and Canada legalized and issued licenses to allow industrial hemp production (Fike, 2016). Even though hemp is still classified as a 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 36 states have enacted hemp bills, and those of 19 states are allowed to grow and cultivate hemp in 2017.[36] Compared to other cultivating states, the states of Colorado and Kentucky are the two predominate with planted acres of 9,700 and 3,100 acres in 2017, respectively. Colorado and Kentucky acres expanded by 64% and 23% from 2016 to 2017.

3.3.3 Current Retail Sales of U.S. Hemp Products

Nielsen Retail Scanner data provides consumption and accessibility information to allow us to gain a deeper understanding of the U.S. hemp market. The scanner data contains weekly pricing, volume, and store information based on a point-of-sale system with more than 90 participating retail chains in the U.S. Table 3.1 demonstrates the quantity sold for hemp products– granola, nuts, nutrition, and protein–by region in the U.S. from 2008 to 2015. The regions in Table 1 are based on four statistical regions defined by the U.S. Census Bureau: Northeast, Midwest, South, and West. As shown in Table 3.1, the total quantity sold in each category of hemp products is increasing over time regardless of the regions. The sales volume of hemp granola, especially, is much higher than other hemp products, and noticeably about 40% of hemp granola is sold in the West region. This implies there might be many stores selling granola hemp, and consumers might have better accessibility in the West region. For the category of hemp nuts, approximately 33% and

[36] Please see more detail about state hemp legislation at http://www.votehemp.com/PR/PDF/Vote-Hemp-2017-US-Hemp-Crop-Report.pdf

31% of hemp nuts are sold in the Northeast and West regions, respectively. Also, about 65% of hemp nutrition are sold mainly in the West and Northeast regions: approximately 38% in the west and 28% in the Northeast. Compared to other hemp products, hemp protein is sold mainly in the West with 43% and Midwest with 31%. For hemp protein, the sales volume in the South region is steadily decreasing since 2008 and rebounding from 2012. To sum, consumption of most hemp products show an increasing trend from 2008 to 2015 although there are little variations in four different regions and years. Particularly from 2008 to 2015, the amount of hemp granola sold in the Northeast decreased by approximately 48% while hemp protein in the Southeast decreased by 42%.

3.4 Data Description

The consumer panel data started in 2004 and is updated with a 2-year time lag. The database contains information about product purchases 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 detailed information about demographic and geographic information of the panelists, products, product characteristics, retail channels, and market location.

Consumer Panel product data are organized based on the hierarchy as follow: departments, product groups, product modules, and UPC codes. In the first step, we employ a searching index function based on a string of characters that include “hemp” to identify the product hierarchy. 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. In the second step, we narrow the product groups down to the next hierarchy, which is product modules, to identify whether there are any missing information or irrelevant products that are associated with the four product groups. In this step, this study excludes the product modules if there are no or only a few observations to represent the product groups identified in the first step. In the third step, we collect all households from the Nielsen Consumer Panel data and limit the panelists to four main product categories: granola, nut, nutritional supplement, and protein supplements. In the final step, we exclude households based on the store code, which is uniquely assigned for each household. It is due to the fact that some households may not be accessible to buy hemp products if stores do not sell products that contain hemp. In this case, we are not able to identify and differentiate factors that make consumers more likely to buy products that contain hemp than conventional products. Through these steps, we explicitly classify the hemp consumers and estimate the probability of purchasing hemp products and the impact of characteristics of households on total hemp expenditures. Table 3.2 shows the number of observations for each product with the proportion of hemp products. For nuts, for example, there are total 15,241 households who consume nuts from 2008 to 2015, and 11.20 percent of them consume hemp nuts.

The demographic and socioeconomic characteristics, especially, education level, age, race, and ethnicity in Nielsen's consumer data contains both the male and the female head of households. Since the head of the household is either male or female head, this study mainly uses female demographic information by assuming that females make the majority of grocery shopping. This assumption is consistently applied to previous studies such as Dettmann (2008) and Alviola and Capps (2010) that use Nielsen's consumer data. If the female head of household does not exist, then the head of the household is replaced with the male head of household. Table 3.3 shows the summary statistics of variables used in the analysis.

Many of the demographic and socioeconomic variables in Nielsen's consumer data are classified into many different group categories. This study reclassifies some of them to be 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 the household head is reclassified from nine categories into three categories: less than 40 years, between 40 and 64 years, and over 64 years. Finally, the education of the household head is reclassified from six into four categories: high school or less, some college, college graduate, and post-collegiate.

In addition to the demographic and socioeconomic characteristics, this study incorporates a new variable called hemp legislation if the state has enacted any hemp legislation that allows for the production of hemp in the state. Therefore, this study hypothesizes that households in states where hemp bills and resolutions introduced are more likely to be exposed to hemp products compared to households in other states.

The variable Hemp is used as a dependent variable for the probit model and is defined as 1 to represent the purchase of hemp product and 0 otherwise. The sample of households purchasing hemp products of granola, nuts, nutrition, and protein are approximately 24%, 11%, 2%, and 13%, respectively from 2008 to 2015. The proportions of household income with low, medium and high levels across all products are roughly 10%, 37%, and 50%, respectively. On average, the household sizes are roughly 2.4, 70% are married, and more than 50% of the households are between 40 and 64 years old across the products. For other demographic characteristics, on average across the products, more than 50% of the head of households are employed, and 80% plus have at least some college. This study also includes race and the sample is approximately classified as white, black, Asian, and other races with 80%, 8%, 5%, and 6%, respectively. Additionally, about 8% of the sample are classified as Hispanic. Finally, this study includes four regional dummies as Midwest, South, West, and East, and the majority of the households, on average across the products, are in the West (about 37%), followed by the South, Midwest, and East.[37] Even though this study includes year dummies to avoid and control for potential heterogeneity across years, we do not report them in Table 3.3.

3.5 Empirical Methodology

This paper employs the Heckman sample selection approach (also called a two-step model) developed by Heckman (1979) to correct for sample selection bias from non-randomly selected samples. Therefore, this study estimate the likelihood of market participation and consumption level. The Heckman selection model is different from other approaches such as Tobit model and Cragg's model (also known as the hurdle

[37] In the Nielsen Consumer Panel, the regions are originally classified with 9 different regions However, we reclassified 9 regions with 4 major regional distinctions: East includes New England and Middle Atlantic, Midwest includes East North Central and West North Central, South includes South Atlantic, East South Central, and West South Central, and finally West includes Mountain and Pacific.

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 as follows.

3.5.1 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:

$$Pr(z_i = 1) = \Phi(W_i \gamma) \tag{1}$$

where z_i is an indicator that takes on the value of 1 if the household *i* buys hemp product and 0 otherwise, Φ is the standard normal cumulative distribution function, and W_i is the vector of explanatory variables for the decision to buy hemp 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 following:

$$IMR = \hat{\lambda}_i(W_i\hat{\gamma}) = \frac{\phi(W_i\hat{\gamma})}{\Phi(W_i\hat{\gamma})} \tag{2}$$

where $\phi(W_i\hat{\gamma})$ is the estimated probability density function (*pdf*), and $\Phi(W_i\hat{\gamma})$ is the cumulative density function (*cdf*). The calculated *IMR* indicates the probability that the household *i* decided to buy hemp 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).

3.5.2 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 products influence the total expenditure. Therefore, the regression model for the selected sample in the second stage is mathematically formed as

$$E(Y_i|z_i = 1) = X_i\beta + \alpha\hat{\lambda}_i(W_i\hat{\gamma}) + v_i \tag{3}$$

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

3.5.3 Marginal Effects of the Heckman Model

The 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

$$\widehat{ME}_{ij} = \frac{\partial E(Y_i|z_i = 1)}{\partial X_i} = \beta_j + \alpha \frac{\partial IMR_i}{\partial X_{ij}} \tag{4}$$

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 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

$$\widehat{ME}_{ij} = \hat{\beta}_j - \hat{\alpha}\hat{\gamma}\left(W_i\hat{\gamma}\hat{\lambda}_i + (\hat{\lambda}_i)^2\right) \tag{5}$$

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 products. Saha, et al. (1997) and Alviola and Capps (2010) ague $\widehat{ME}_{ij} \neq \hat{\beta}_j$ in general, but $\widehat{ME}_{ij} = \hat{\beta}_j$ if and only if $\hat{\alpha} = 0$, implying covariance of two error terms between first- and second-stage equations are equal to zero. Since this case is not common, and the ME is different across the observation (i.e., observation dependent), this paper evaluates the ME at the sample mean as follows:

$$\widehat{ME}_{ij}|_{\boldsymbol{sample\ mean}} = \hat{\beta}_j - \hat{\lambda}\hat{\gamma}_j\left((\bar{W}\hat{\gamma})\bar{\hat{\lambda}} + \bar{\hat{\lambda}}^2\right) \tag{6}$$

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.

3.6 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, employment, and hemp state.[38] The mathematical expression of the probit model for the decision to purchase hemp products is written as follows:

$$\begin{aligned} Pr(z_i = 1) = \gamma_0 + \gamma_1 M_{Income} + \gamma_2 H_{Income} + \gamma_3 Age2 + \gamma_4 Age3 \\ + \gamma_5 HHSize + \gamma_6 Married + \gamma_7 Edu2 + \gamma_8 Edu3 + \gamma_9 Edu4 \\ + \gamma_{10} White + \gamma_{11} Black + \gamma_{12} Asian + \gamma_{13} Hispanic \\ + \gamma_{14} Employ + \gamma_{15} Hemp\_State + \epsilon_i \end{aligned} \tag{7}$$

[38] We test multicollinearity between the variables based on Variance Inflation Factor (VIF) across all different categories of hemp products, and we find there is no strong evidence of multicollinearity.

A description of the variable names in the equation above is based on Table 3.3 with associated descriptive statistics. The reference category for each dummy variable scheme is excluded and reported with an asterisk symbol in Table 3.3. Regional and year dummies are also included in the estimation but are not reported in the equation above.[39]

$$\begin{aligned}\ln(Expenditure) \\ &= \beta_0 + \beta_1 M\_Income + \beta_2 H\_Income + \beta_3 Age2 + \beta_4 Age3 \\ &+ \beta_5 HHSize + \ \beta_6 Married + \beta_7 Edu2 + \beta_8 Edu3 \qquad (8) \\ &+ \beta_9 Edu4 + \beta_{10} White + \beta_{11} Black + + \beta_{12} Asian \\ &+ \beta_{13} Hispanic + \beta_{14} Employ + \lambda_1 IMR + u_i\end{aligned}$$

For the dependent variable, this study uses aggregated monthly expenditure, and we transform the dependent variable into logarithm form. Econometrically, more outliers in the linear dependent variable reflect high variance and result in more risk of heteroskedasticity. Also, the distribution of consumption data is commonly highly skewed (Zhang, et al., 2008), and estimators might be inconsistent with a dependent variable without transformation (Newman, et al., 2003). According to Newhouse (1987), Wagner and Hanna (1983), Zhang, et al. (2008), the transformation of the natural logarithm will control for positively skewed expenditures. Figure 3.2 shows the histograms for hemp products and the distributions of the values are positively skewed across all categories, implying that most consumers are spending small amounts of money to purchase hemp products.[40] On the other hand, the distributions of the expenditures in the natural logarithm

[39] For the time dummies, this paper employs year fixed effects instead of monthly fixed effects due to the fact we find there is no significant variation among months.

[40] For both Figures 3 and 4, we only use the positive value of the expenditures that head of household who purchase hemp products.

of hemp products are more normally distributed as shown in Figure 3.3. Therefore, we use the logarithm dependent variable to reduce problems resulting from a non-normal distribution. In the second stage estimation, we exclude the variable of hemp state in that it is not atypical in Heckman selection model. Also, the variable of *IMR* calculated from the probit model is included to test the selection bias.

3.7 Empirical Results

3.7.1 First-Stage Estimation

The results of the first stage probit model for four different categories are reported in Table 3.4 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 3.4 since the magnitude of the coefficients does not provide direct interpretation. By looking at the marginal effects in Table 3.4, households with higher income are more likely to consume hemp granola and nuts, relative to low-income categories with less than $30,000. Older households are less likely to consume all hemp products except hemp protein compared to the younger households who are less than 40-year-old, indicating young households are more likely consumers of hemp products. For the education level, we find that most of the categories of hemp products except hemp granola are more likely to be consumed as education increases. Also, we find that significant regional effects on the probability of buying hemp products, but the regional effects vary across the categories of hemp products. This finding suggests that consumers may have different preferences for hemp products across the regions regardless of the quantity sold. For example, the likelihood of buying hemp granola is less in the South but more in the Midwest and West regions compared to the Northeast region even though quantity sold for hemp granola in the South is about 13% higher compared to Northeast based on Table 3.1. States that have enacted industrial hemp legislation are more likely to consume hemp granola but less likely to consume hemp nuts relative to no hemp legislation states. This could be a function of no processing of hemp and minimal advertising of hemp in states without hemp legislation.

After the estimation of the probit model, the prediction success is evaluated to assess the usefulness of the probit model as suggested by other studies such as (Alviola and Capps, 2010, Capps, et al., 1999, Park and Davis, 2001). Table 3.5 shows the goodness of fit measures from the probit model for all four categories. To generate the classification statistics, especially the percentage of correct predictions, we employ different cut off values for each category rather than the default value of 0.5. This is used because the classification of households who purchased hemp products are incorrectly classified if the default value is used instead of market penetration (Alviola and Capps, 2010). Therefore, the cut off value represents the market penetration that is the proportion of the households who purchase hemp products. As shown in Table 3.5, the percentage of correct predictions between hemp products and non-hemp products of granola, nuts, nutrition, and proteins are 61.97%, 53.65%, 59.77%, and 61.82%, respectively. Based on the sensitivity in Table 3.5, our models correctly predict the decision to buy hemp-products of granola, nuts, nutrition, and protein: approximately 63%, 72%, 67%, and 63%, respectively. Regarding specificity, the decision to purchase non-hemp products of granola, nuts, nutrition, and proteins are correctly predicted approximately 64%, 51%, 59%, and 61%, respectively.

3.7.2 Second-Stage Estimation

The results of the second stage estimation are reported in Table 3.6. 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 granola, 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 the use of Heckman selection model is justified. In Table 3.6, we also reported the marginal effects that are evaluated at the mean due to the observation dependent. Also, the marginal effect used in this study is the partial effect on the truncated mean. The marginal effect is calculated based on consumers who have an observed value by excluding consumers who do not purchase hemp products. For the second stage estimation, once a decision to buy hemp products has been made by households (i.e., hemp buyers), higher income group households are positively associated with total expenditure but only for hemp granola. If household's income is above \$70.000, the total expenditure of hemp granola increases by 9.8% compared to the household whose income is less than \$30,000, *ceteris paribus*. Total expenditure for the hemp nuts and nutrition are positively associated with households in higher age groups. To be specific, on average, the total expenditure of hemp nuts increases by 2.5% while the total expenditure of hemp protein decreases by 19.2% if household's age is above 74 compared to the household whose age is less than 40. This finding suggests that the younger age group may be looking for healthier sources of protein than the older age group. Across all categories of hemp products, we find higher education level is not statistically related to the total expenditure across most categories. For the different regions, households in the South and West regions consume less for hemp granola compared to the households in the East region whereas households in the Midwest and South regions consume more hemp nutrition and protein, relative to households in the East region. The findings can be explained by two potential reasons. First, households in South and West regions have less accessibility to stores and availability of products to purchase hemp granola compared to the households in the East region. Likewise, there might be more stores that carry hemp nutrition and protein in the Midwest and South regions compared to the East region. Second, households in the Midwest and South regions might have a stronger perception of hemp products as nutritious and protein-rich than households in the East region.

3.8 Concluding Remarks

Industrial hemp as a variety of the Cannabis sativa plant species has received a great deal of interest in the last two decades since there are many benefits in environmental, production, and health. The passage of the 2014 Farm Bill only accelerated the interest in this crop and its potential. In global markets, industrial hemp is an agricultural crop used for textiles, automotive paneling, furniture, food, personal care, construction, paper, etc. Then on December 2018, the 2018 Farm Bill was passed and approved by the U.S. Congress to legalize the production of industrial hemp. Ratification of this legislation would open up the opportunity for commercial hemp production and increasing the supply of hemp available in the U.S. market.

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. This study investigates the critical sociodemographic factors that are associated with increasing hemp consumption and measures their effects on total expenditure in the U.S. by utilizing Nielsen's consumer panel data from 2008 to 2015. By analyzing the retail data, a more objective view of the consumer profile is identified for this developing industry. Knowing this consumer profile, therefore, can contribute to the viability of the hemp products market in the U.S.

By employing the Heckman selection model, this study finds that sociodemographic characteristics especially income, age, and education play an important role in purchasing and explaining the demand for different categories of hemp products. To be specific, higher income is positively associated with the probability and consumption level of hemp products. The role of the age of household head is mixed with respect to consumption decisions and consumption level across the products: a negative and significant effect on the probability of buying hemp products except for hemp protein, whereas a positive and significant effect on total expenditure of hemp nuts and nutrition once households make the decision to buy hemp products. In most of the cases, households with higher education are more likely to buy hemp products, and those households are associated with significantly higher levels of the consumption except for hemp nuts. To understand the hemp market in the U.S., these findings will provide insights into a more targeted marketing strategy for hemp industries to attract new consumers and increase more sales from current consumers. Many different markets such as hemp seed, hemp fiber, and hemp CBD can be derived from the hemp industry since more than fifty thousand uses are produced from hemp. Hemp products used in this study are made of hemp seeds; however, our findings show that hemp seeds market could be segmented based on the forms: hemp cereal, hemp nuts, hemp nutrition, and hemp protein.

Industrial hemp is recently removed from the Schedule 1 narcotic and is legal to produce in the U.S. according to the 2018 Farm Bill. As of 2018, 40 plus states already have hemp legislation in place that allows for the production and processing of industrial hemp within the state. Based on our best knowledge, there is no empirical study related to 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 U.S. As the industry continues to move forward, findings in this study may also open the door to create a business and marketing plans that allow to create goals and strategies to marketers, retailers, and other stakeholders. Not only will this manuscript contribute to the industrial hemp literature, but it has the potential to generate significant discussion. Little is known about modern 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.

3.9 Table and Figures

Table 3.1 The Quantity of Hemp Products Sold by Region in the U.S.

Products	Regions	2008	2009	2010	2011	2012	2013	2014	2015
Granola	Northeast	27,469	17,054	14,508	19,855	23,070	19,023	13,012	14,171
	Midwest	36,299	32,427	37,898	42,753	45,648	41,713	37,471	36,763
	South	32,387	32,201	34,556	41,673	45,095	42,724	38,065	46,816
	West	34,787	37,400	60,834	73,067	75,967	65,929	68,496	77,001
	Total	**130,942**	**119,082**	**147,796**	**177,348**	**189,780**	**169,389**	**157,044**	**174,751**
Nuts	Northeast	467	577	970	1,873	6,001	18,687	23,220	25,808
	Midwest	166	349	751	1,189	2,902	7,607	15,915	19,944
	South	222	253	426	613	998	3,559	10,271	20,181
	West	-	79	214	322	618	1,708	26,944	42,806
	Total	**855**	**1,258**	**2,361**	**3,997**	**10,519**	**31,561**	**76,350**	**108,739**
Nutrition	Northeast	-	9	26	198	863	10,470	21,753	20,405
	Midwest	45	253	94	150	1,523	6,305	13,531	15,703
	South	104	504	32	-	63	553	12,958	14,516
	West	4,073	3,312	2,959	6,416	9,303	12,216	19,194	16,999
	Total	**4,222**	**4,078**	**3,111**	**6,764**	**11,752**	**29,544**	**67,436**	**67,623**
Protein	Northeast	-	54	47	32	65	185	314	8,149
	Midwest	253	1,540	1,018	1,367	2,230	3,256	3,403	5,027
	South	1,029	3,023	187	153	170	270	276	596
	West	660	6,106	4,806	3,235	3,813	4,845	967	494
	Total	**1,942**	**10,723**	**6,058**	**4,787**	**6,278**	**8,556**	**4,960**	**14,266**

Table 3.2 Number of Observations for Each Product with Proportion of Hemp Product

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)
Nutrition	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)
Total	3,886	3,772	2,093	2,877	4,842	7,710	8,941	8,228	42,349
	(6.56)	(7.41)	(13.09)	(12.00)	(10.55)	(10.96)	(11.49)	(11.80)	(10.66)

Notes: Observations in Table 3.2 represent the total number of observations for each product type, regardless of whether it contains a hemp product. Parentheses represent the proportion of these observations for which a hemp product is included in the ingredients.

Table 3.3 Definitions and Summary Statistics of Variables Used in the Analysis

Variable	Description
Hemp Exp	Total monthly expenditure for Hemp product in log
Hemp	=1 if HH consume Hemp by product
Low Income*	=1 if HH income is less than $30,000
Median Income	=1 if HH income is between $30,000 and $70,000
High Income	=1 if HH income is above $70,000
Age1*	=1 if HH age is less than 40
Age2	=1 if HH age is between 40 and 64
Age3	=1 if HH age is above 64
HH Size	Size of Households
Married	=1 if HH married
Edu1*	=1 if HH education is High School or less
Edu2	=1 if HH education is Some College
Edu3	=1 if HH education is College Graduate
Edu4	=1 if HH education is Post Collegiate
White	=1 if HH is White
Black	=1 if HH is African American (Black)
Asian	=1 if HH is Asian
Other Race*	=1 if HH is other races
Hispanic	=1 if HH is Hispanic
Employ	=1 if HH is employed
Hemp State	=1 if HH is living in State with Hemp Legislation
Midwest	=1 if HH is living Midwest region
South	=1 if HH is living South region
West	=1 if HH is living West region
East*	=1 if HH is living East region

Notes: 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. A variable with an asterisk symbol represents a reference (base) category.

Table 3.3 (Continued)

	Granola		Nuts		Nutrition		Protein	
Variable	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Hemp Exp	1.40	0.48	2.45	0.54	2.44	0.47	2.69	0.39
Hemp	0.24	0.43	0.11	0.32	0.02	0.15	0.13	0.33
Low Income*	0.10	0.30	0.14	0.35	0.14	0.35	0.11	0.31
Median Income	0.37	0.48	0.39	0.49	0.39	0.49	0.33	0.47
High Income	0.53	0.50	0.47	0.50	0.47	0.50	0.56	0.50
Age1*	0.14	0.34	0.09	0.28	0.07	0.26	0.13	0.34
Age2	0.65	0.48	0.65	0.48	0.63	0.48	0.67	0.47
Age3	0.21	0.41	0.26	0.44	0.29	0.46	0.20	0.40
HH Size	2.52	1.26	2.39	1.19	2.28	1.16	2.45	1.29
Married	0.73	0.44	0.71	0.46	0.67	0.47	0.68	0.47
Edu1*	0.15	0.36	0.20	0.40	0.18	0.39	0.13	0.34
Edu2	0.27	0.44	0.30	0.46	0.31	0.46	0.32	0.47
Edu3	0.36	0.48	0.34	0.47	0.35	0.48	0.36	0.48
Edu4	0.22	0.41	0.16	0.36	0.16	0.36	0.18	0.39
White	0.84	0.37	0.81	0.40	0.81	0.40	0.78	0.42
Black	0.06	0.23	0.09	0.29	0.09	0.28	0.11	0.31
Asian	0.05	0.21	0.05	0.22	0.05	0.23	0.05	0.21
Other Race*	0.06	0.23	0.06	0.23	0.05	0.23	0.07	0.25
Hispanic	0.08	0.27	0.08	0.27	0.07	0.26	0.08	0.27
Employ	0.64	0.48	0.57	0.50	0.56	0.50	0.64	0.48
Hemp State	0.37	0.48	0.40	0.49	0.40	0.49	0.48	0.50
Midwest	0.20	0.44	0.21	0.41	0.15	0.36	0.17	0.38
South	0.26	0.44	0.31	0.46	0.31	0.46	0.29	0.45
West	0.35	0.48	0.32	0.47	0.41	0.49	0.41	0.49
East*	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.

Table 3.4 First Stage Probit Estimation Results

	Granola (N=5,959)			Nuts (N=15,233)		
Variable	Coef		M.E	Coef		M.E
M_Income	0.154	**	0.047	0.151	***	0.026
	(0.069)		(0.021)	(0.047)		(0.008)
H_Income	0.129	*	0.038	0.167	***	0.029
	(0.072)		(0.021)	(0.050)		(0.009)
Age2	-0.199	***	-0.061	-0.083		-0.014
	(0.054)		(0.017)	(0.051)		(0.009)
Age3	-0.294	***	-0.082	-0.141	**	-0.023
	(0.069)		(0.018)	(0.059)		(0.009)
HH Size	-0.084	***	-0.025	-0.021		-0.004
	(0.018)		(0.005)	(0.015)		(0.002)
Married	-0.066		-0.02	-0.157	***	-0.028
	(0.050)		(0.015)	(0.037)		(0.007)
Edu2	-0.053		-0.016	0.121	***	0.021
	(0.060)		(0.018)	(0.042)		(0.008)
Edu3	-0.039		-0.012	0.1	**	0.017
	(0.059)		(0.017)	(0.042)		(0.007)
Edu4	0.029		0.009	0.184	***	0.034
	(0.065)		(0.020)	(0.049)		(0.010)
Employed	0.134	***	0.04	0.015		0.003
	(0.043)		(0.013)	(0.031)		(0.005)
White	-0.259	***	-0.082	-0.206	***	-0.038
	(0.084)		(0.028)	(0.061)		(0.012)
Black	-0.013		-0.004	-0.281	***	-0.041
	(0.114)		(0.034)	(0.076)		(0.009)
Asian	-0.504	***	-0.123	-0.321	***	-0.045
	(0.123)		(0.024)	(0.088)		(0.010)
Hispanic	-0.011		-0.003	0.147	***	0.027
	(0.073)		(0.022)	(0.053)		(0.010)
Midwest	0.159	***	0.049	-0.097	*	-0.016
	(0.059)		(0.019)	(0.051)		(0.008)
South	-0.202	***	-0.058	0.26	***	0.047
	(0.059)		(0.016)	(0.046)		(0.009)
West	0.268	***	0.082	0.269	***	0.049
	(0.053)		(0.017)	(0.044)		(0.008)
Hemp State	0.086	*	0.026	-0.055	*	-0.009
	(0.044)		(0.013)	(0.032)		(0.005)
Constant	-0.302	**		-1.94	***	
	(0.145)			(0.125)		
Log Likelihood			-3054.221			-5035.573
McFadden R^2			0.069			0.058

Table 3.4 (Continued)

	Nutrition (N=18,871)			Protein (N=2,263)		
Variable	Coef		M.E	Coef		M.E
M_Income	0.041		0.002	0.035		0.007
	(0.065)		(0.003)	(0.128)		(0.025)
H_Income	-0.107		-0.005	-0.077		-0.015
	(0.070)		(0.003)	(0.126)		(0.025)
Age2	-0.298	***	-0.016	0.084		0.016
	(0.066)		(0.004)	(0.107)		(0.020)
Age3	-0.403	***	-0.016	-0.113		-0.021
	(0.078)		(0.003)	(0.148)		(0.026)
HH Size	0.005		0.0002	-0.03		-0.006
	(0.021)		(0.001)	(0.034)		(0.007)
Married	0.098	*	0.004	0.063		0.012
	(0.054)		(0.002)	(0.096)		(0.018)
Edu2	0.163	***	0.008	0.241	**	0.049
	(0.063)		(0.003)	(0.117)		(0.025)
Edu3	0.178	***	0.009	0.237	**	0.048
	(0.064)		(0.003)	(0.119)		(0.025)
Edu4	0.006		0.0003	-0.035		-0.007
	(0.081)		(0.004)	(0.138)		(0.026)
Employed	-0.006		-0.0003	0.159	*	0.03
	(0.045)		(0.002)	(0.086)		(0.016)
White	-0.024		-0.001	0.095		0.018
	(0.098)		(0.005)	(0.158)		(0.029)
Black	0.15		0.008	0.019		0.004
	(0.114)		(0.007)	(0.195)		(0.038)
Asian	-0.283		-0.01	0.167		0.035
	(0.149)		(0.004)	(0.223)		(0.051)
Hispanic	-0.098		-0.004	0.031		0.006
	(0.089)		(0.004)	(0.148)		(0.030)
Midwest	-0.192	***	-0.008	0.147		0.03
	(0.067)		(0.002)	(0.124)		(0.027)
South	-0.321	***	-0.014	0.092		0.018
	(0.060)		(0.002)	(0.115)		(0.023)
West	-0.307	***	-0.014	-0.279	**	-0.052
	(0.057)		(0.002)	(0.118)		(0.021)
Hemp State	0.001		0.0001	-0.027		-0.005
	(0.057)		(0.003)	(0.082)		(0.016)
Constant	-2.09	***		-1.068	***	
	(0.189)		(0.002)	(0.297)		
Log Likelihood			-2027.504			-812.334
McFadden R^2			0.046			0.058

Notes: Significance levels are indicated by ***, **, * for 1, 5, and 10 percent significance level, respectively. Robust standard errors are reported in parentheses.

Table 3.5 The Goodness of Fit Measures from the Probit Model

Categories	Sensitivity (%)	Specificity (%)	Cutoff Value	% of Correct Predictions
Granola	63.77	64.41	0.239	61.97%
Nuts	72.47	51.28	0.112	53.65%
Nutrition	67.11	59.6	0.024	59.77%
Protein	63.19	61.62	0.127	61.82%

Notes: Sensitivity represents the percentage of correctly predicting the choice of hemp products, whereas specificity represent the percentage of correctly predicting the choice of choosing non-hemp products

Table 3.6 Second Stage Estimation Results

	Granola			Nuts		
Variable	Coef		M.E	Coef		M.E
M_Income	0.080	*	0.085	-0.039		0.058
	(0.044)		(0.044)	(0.053)		(0.044)
H_Income	0.094	**	0.098	0.004		0.106
	(0.045)		(0.046)	(0.055)		(0.046)
Age2	-0.110	***	-0.117	0.091	*	0.042
	(0.040)		(0.040)	(0.053)		(0.042)
Age3	-0.010		-0.021	0.113	*	0.025
	(0.049)		(0.048)	(0.066)		(0.054)
HH Size	-0.064	***	-0.067	0.028	*	0.014
	(0.014)		(0.014)	(0.015)		(0.012)
Married	0.096	***	0.094	0.049		-0.055
	(0.033)		(0.033)	(0.043)		(0.036)
Edu2	0.063		0.061	-0.072		0.008
	(0.042)		(0.042)	(0.048)		(0.038)
Edu3	0.003		0.002	-0.070		-0.007
	(0.040)		(0.040)	(0.046)		(0.037)
Edu4	-0.029		-0.028	-0.117	**	0.006
	(0.044)		(0.043)	(0.056)		(0.044)
Employed	0.072	***	0.077	-0.061		-0.052
	(0.027)		(0.027)	(0.037)		(0.030)
White	-0.071		-0.081	0.158	**	0.039
	(0.055)		(0.055)	(0.064)		(0.052)
Black	-0.119	*	-0.119	0.164	**	0.000
	(0.065)		(0.065)	(0.083)		(0.067)
Asian	-0.135		-0.154	0.186	**	-0.014
	(0.084)		(0.083)	(0.092)		(0.073)
Hispanic	-0.144	***	-0.145	-0.040		0.056
	(0.049)		(0.049)	(0.060)		(0.049)
Midwest	0.029		0.035	0.064		-0.001
	(0.045)		(0.044)	(0.060)		(0.049)
South	-0.105	**	-0.113	-0.083		0.073
	(0.041)		(0.041)	(0.053)		(0.042)
West	-0.076	*	-0.066	-0.054		0.115
	(0.039)		(0.038)	(0.052)		(0.042)
Lambda	-0.046	*	–	-0.765	***	–
	(0.026)		–	(0.047)		–
Constant	1.493	***	–	3.698	***	–
	(0.097)		–	(0.195)		–
Log Likelihood			-3,976.47			-6,340.95
Censored			4,532			13,526
Uncensored			1,427			1,707
Observations			5,959			15,233

Table 3.6 (Continued)

	Nutrition			Protein		
Variable	Coef		M.E	Coef		M.E
M_Income	0.102		0.115	0.089		0.073
	(0.070)		(0.066)	(0.095)		(0.058)
H_Income	0.119		0.087	0.056		0.083
	(0.078)		(0.074)	(0.097)		(0.061)
Age2	0.226	***	0.136	-0.027		-0.073
	(0.071)		(0.060)	(0.079)		(0.048)
Age3	0.206	**	0.083	-0.247	**	-0.192
	(0.091)		(0.075)	(0.106)		(0.063)
HH Size	-0.035		-0.034	-0.013		0.001
	(0.023)		(0.022)	(0.022)		(0.015)
Married	0.077		0.107	-0.036		-0.063
	(0.060)		(0.058)	(0.071)		(0.045)
Edu2	-0.072		-0.023	0.077		-0.023
	(0.073)		(0.066)	(0.086)		(0.055)
Edu3	-0.031		0.023	0.062		-0.041
	(0.071)		(0.065)	(0.089)		(0.057)
Edu4	0.002		0.004	-0.114		-0.064
	(0.085)		(0.081)	(0.102)		(0.065)
Employed	-0.042		-0.044	0.093		0.002
	(0.050)		(0.048)	(0.064)		(0.039)
White	0.06		0.053	0.038		-0.034
	(0.090)		(0.088)	(0.113)		(0.056)
Black	0.122		0.167	0.029		-0.01
	(0.112)		(0.111)	(0.144)		(0.085)
Asian	0.181		0.096	-0.091		-0.179
	(0.175)		(0.166)	(0.161)		(0.098)
Hispanic	0.183	*	0.153	-0.014		-0.016
	(0.099)		(0.095)	(0.107)		(0.056)
Midwest	0.254	***	0.195	0.237	***	0.148
	(0.073)		(0.074)	(0.088)		(0.052)
South	0.328	***	0.229	0.247	***	0.173
	(0.073)		(0.071)	(0.081)		(0.049)
West	0.391	***	0.297	0.067		0.184
	(0.074)		(0.071)	(0.092)		(0.059)
Lambda	-0.34	*	–	0.657	***	–
	(0.194)		–	(0.059)		–
Constant	3.101	***	–	1.811		–
	(0.530)		–	(0.209)		–
Log Likelihood			-2,278.85			-894.196
Censored			18,421			1,975
Uncensored			450			288
Observations			18,871			2,263

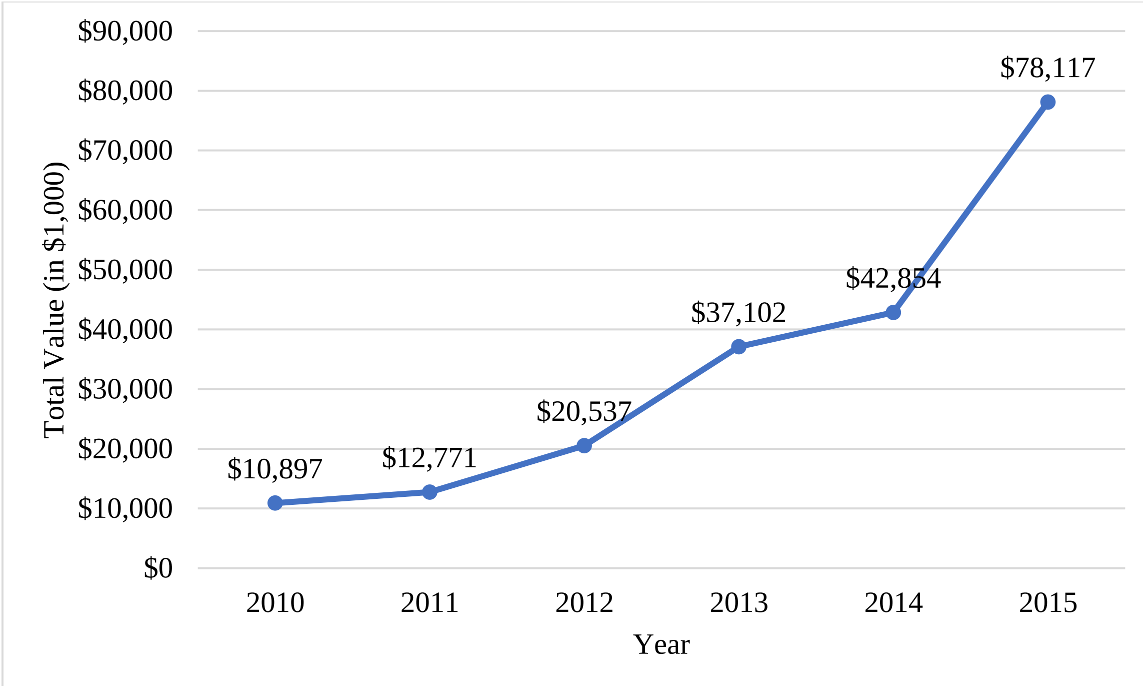
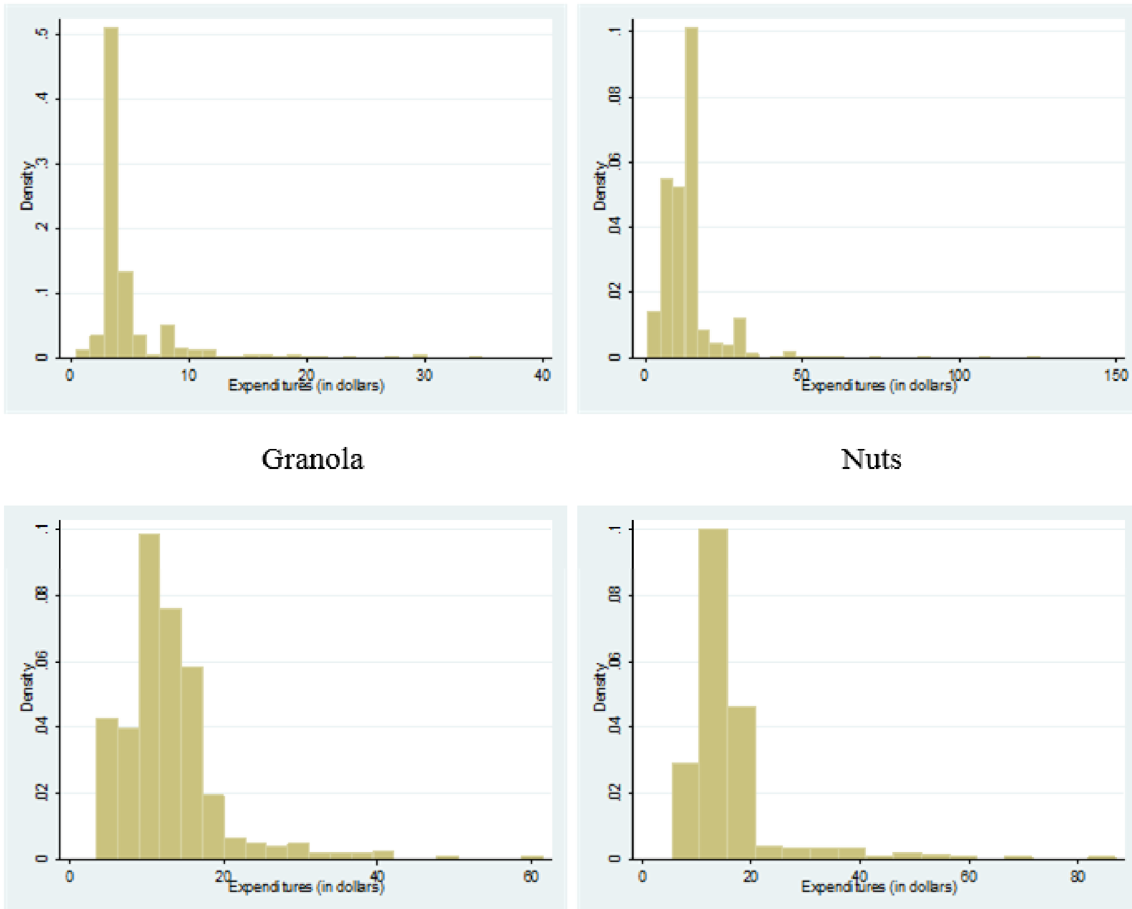


Figure 3.1 Total Value of U.S. Hemp Imports, 2010-2015

Notes: Main source of the 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. Please see more detail information on U.S. hemp import at https://fas.org/sgp/crs/misc/RL32725.pdf



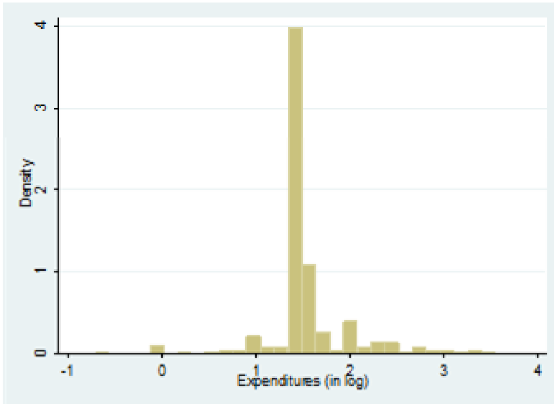
Granola

Nuts

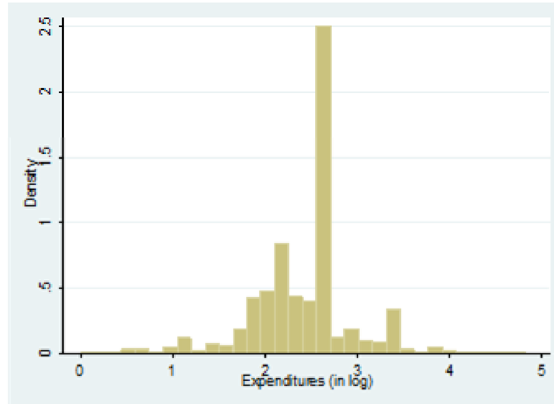
Nutrition

Protein

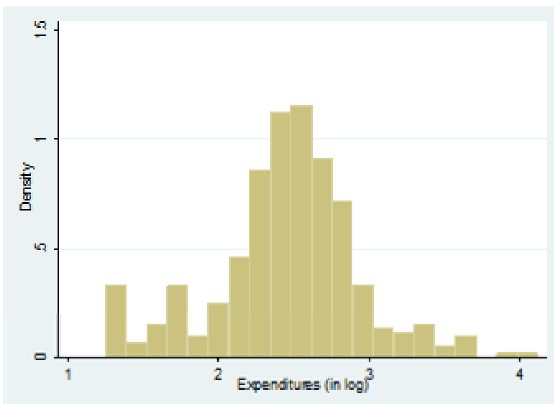
Figure 3.2 Distributions of Hemp Products Expenditures in Original Scale



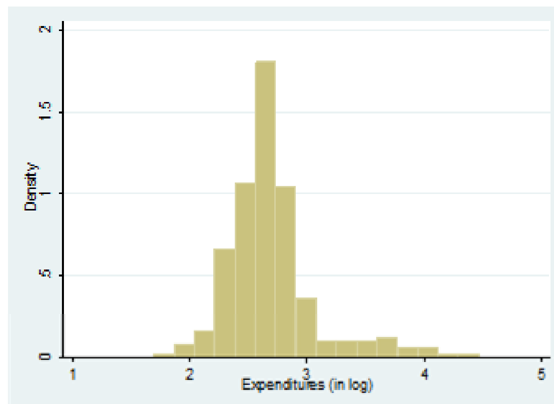
Granola



Nuts



Nutrition



Protein

Figure 3.3 Distributions of Hemp Products Expenditures in Original Scale

CHAPTER 4. FACTORS AFFECTING HETEROGENEOUS AGRICULTURAL LAND: THE CASE OF KENTUCKY

4.1 Abstract

This study investigates the factors that affect agricultural land values by proposing a new rich dataset, Zillow Transaction and Assessment Data (ZTRAX) provided by Zillow from 2009 to 2014. This study also examines whether the National Commodity Crop Productivity Index (NCCPI) could be a good indicator of land values or not by comparing two different regression models between county-level cash rent and parcel-level NCCPI. Finally, this study incorporates flexible functional forms to test the parcel size and land values relations. Findings show that factors influencing agricultural land values in states with heterogeneous agricultural lands such as Kentucky are not different from other states with relatively homogeneous agricultural lands. This study also provides suggestive evidence that there is a non-linear relationship between parcel size and land values. Furthermore, we find that a disaggregated NCCPI at parcel-level could be considered an acceptable indicator to estimate agricultural values compared to an aggregated cash rent at county-level.

4.2 Introduction

Farmland is the source of equity and primary input for most farms; in addition, farmland values are an essential indicator to explain the financial well-being of the farm sector since farmland accounts approximately more than 82 percent of farm-sector assets in 2016 (Burns, et al., 2018). In this regard, not only farmland values but also the factors affecting farmland valuation have been received considerable interest and been the subject of a great deal of economic research (Nickerson, et al., 2012). Abundant research in the literature has examined the factors influencing farmland values and shown that a complex set of factors determines farmland values such as environmental amenities (Bastian, et al., 2002, Borchers, et al., 2014, Wasson, et al., 2013), Urban influence (Delbecq, et al., 2014, Guiling, et al., 2009, Livanis, et al., 2006, Zhang and Nickerson, 2015), Potential land development (Plantinga, et al., 2002), decoupled payment (Ifft, et al., 2015), and wind power facilities (Heintzelman and Tuttle, 2012).

None of the previous studies on land values in the U.S., however, considers the direction and the size of the effect of land size even though some of the studies include the land size as a control variable. Most of the previous studies such as (Borchers, et al., 2014, Delbecq, et al., 2014, Huang, et al., 2006, Zhang and Nickerson, 2015) find that the size of the land parcel has a significantly negative impact on land values. Brorsen, et al. (2015) specifically focus on the relationship between land value and parcel size, and they find that there is an inverse relationship between parcel size and per acre prices of agricultural land. In other words, increasing parcel size leads to a decrease in agricultural land value. A recent working paper by Ritter et al (2019) argues that empirical analysis with hedonic price models is somewhat unclear about the direction and the effect of size. They explain that the ambiguousness regarding the direction and the size effects are attributable to the economies of scale related to farm machinery and management, and partially fixed transaction costs in land values. In addition, the farmland values could be over- or underestimated with a single estimated coefficient of the parcel size due to the large variation that is associated with the parcel size (Ritter et al., 2019). In this respect, our study fills the knowledge gap between the direction and the effect of parcel size by including different functional forms for the parcel size. We also calculate specific threshold points where the direction and effect of parcel size change. This study hypothesizes that the size and land values relations could be non-linear.

A fundamental assumption, which is common to the land value literature, is that land value is the discounted present value of expected returns from the land. One measure of the return to the land is the cash rents (Robison, et al., 1985). According to the Ricardian Rent Theory, cash rents generally reveal the level of profitability of the land (Ibendahl and

Griffin, 2013). Figure 4.1 shows a graphical relationship between agricultural land value and cash rent in the U.S. from 2008 to 2016. As shown in figure 4.1, there is co-movement between land values and cash rents over time, suggesting there is a strong relationship between them. Knowing the relationship between cash rents and land prices is important because it helps indicate whether cash rents are a cost-effective way of controlling farmland relative to purchasing the land. In this regard, we hypothesize that land values can be explained as a function of cash rents, and there is a positive relationship between them.

Based on our best knowledge, however, cash rents at the parcel or field level is not publicly available and difficult to observe. This motivates our research question whether aggregate county-level data (for example, cash rents used in this study) can be substitutable with disaggregated parcel-level or field-level data (for example, NCCPI used in this study).[41] If the disaggregate characteristic predicts land values relatively better than cash rents, it suggests the disaggregate characteristic can be used as an alternative and appropriate indicator when analyzing land values. The productivity of agricultural land in Ricardian rent theory is explained by the natural fertility of the soil (Blaug, 1997). National Commodity Crop Productivity Index (NCCPI) is a national soil interpretation developed by the Conservation Reserve Program (CRP) of USDA and generated in the National Soil Information System.[42] NCCPI is a national soil interpretation that utilizes soil, landscape, and climate factors not only to assign ratings but also model the response for commodity

---

[41] The disaggregated parcel-level data is obtained by using spatial join tool in ArcGIS from NCCPI, which is a raster (i.e., image) file.

[42] NCCPI has three different submodels (i.e., categories): Corn and Soybeans, Small Grains, and Cotton. NCCPI used in this study represents the Corn and Soybean submodel. The corn and soybeans index is calculated by multiplying ratings from the chemical, water, physical, climate, and landscape. Please see more detail information about NCCPI at https://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/16/nrcs143_020559.pdf

crops such as corn and soybeans, small grains, and cotton (Dobos, et al., 2008, Zhang and Nickerson, 2015). Therefore, NCCPI should be representative of cash rents, as a result, land value also could be a function of commodity crop productivity, NCCPI.

The main objective of this paper is three-fold. First, we investigate the factors that affect agricultural land values by proposing a new rich dataset, Zillow Transaction and Assessment Data (ZTRAX) provided by Zillow from 2009 to 2014. This study focuses only on Kentucky where the agricultural lands are heterogeneous, and we hypothesize that the factors influencing farmland values may not be consistent with the findings in the previous studies on farmland values. Figure 4.2 shows the commodity crop productivity in Kentucky. As shown in figure 4.2, eastern Kentucky has lower productivity compared to western Kentucky. This clearly supports evidence that agricultural lands in Kentucky are heterogeneous. Another reason to focus on one single state is based on Palmquist (2005) that it is appropriate to treat a region or state as a single land market. Second, we examine two different regression models with county-level cash rent and with parcel-level NCCPI to test whether NCCPI could be a good indicator of land values. Finally, this study incorporates flexible functional forms of the parcel size to test the parcel size and land values relations. The remainder of this paper is organized as follows. In the next section, we present a conceptual and empirical framework. The data and variables used in this study are discussed in section 4.4. Section 4.5 discusses the results of the analysis, and section 4.6 provides discussion and conclusions.

4.3 Conceptual Framework and Empirical Model

4.3.1 Conceptual Framework

The value of land based on economic theory should be defined as the net present value of future returns, and most previous studies on land value are influenced by Ricardian theory of rend (David, 1817). The relationship between current farmland values and expected returns in future periods is formally expressed in the capitalization formula, which is the foundation of most farmland valuation studies (Ifft, et al., 2015). The following discussion of the capitalization formula is based on Guiling, et al. (2009). The capitalization formula is expressed:

$$Agricultural\ Land\ Values = \frac{Returns}{Discount\ Rate} \quad (2)$$

The equation (1) can be written as the infinite-horizon present value model:

$$V_{it} = \sum_{s}^{\infty} \frac{E_t(R_{is})}{(1+r)^{s-t}}, where\ s = t, t+1, \ldots \quad (2)$$

where V_{it} is the value of a parcel of land *i* at time *t*, R_{is} represents the returns to land *i* in period *t*, *r* is a constant discount rate, and $E_t(\cdot)$ is the expectations operator given the information available at time *t*. Since expectations are unobservable, it is commonly substitutable with cash rents or imputed returns, which are the measure of observed returns. By taking the logarithm of both sides of equation (2), equation (2) can be re-written as follows:

$$ln(V_{it}) = ln(\sum_{s=t}^{\infty} E_t(R_{is})) - ln(1+r)^{s-t} \quad (3)$$

In the empirical model, $ln(\sum_{s=t}^{\infty} E_t(R_{is})$ is approximated by a linear function of parcel characteristics, and the approximation of $ln(\sum_{s=t}^{\infty} E_t(R_{is})$ is expressed as

$$ln(\sum_{s=t}^{\infty} E_t(R_{is}) = \boldsymbol{X}_{it}\boldsymbol{\beta} + \tau_t + \varepsilon_{it} \tag{4}$$

where $\boldsymbol{X}_{it}$ is a vector of factors affecting returns in parcel i, τ_t is a fixed effect for year that controls unobservable effects that change over time, ε_{it} is variation in farmland values that cannot be explained by the model and assumed to be a normally distributed error term. Substituting equation (4) into (3) results

$$ln(V_{it}) = \left(\tau_t - ln(1+r)^i\right) + \boldsymbol{X}_{it}\boldsymbol{\beta} + \varepsilon_{it} \tag{5}$$

In equation (5), both τ_t and $ln(1+r)^i$ can be captured by fixed effect of δ_t that controls unobservable heterogeneity of interest rates over time. With this, an estimable reduced form model from the equation (5) is expressed

$$ln(V_{it}) = \boldsymbol{X}_{it}\boldsymbol{\beta} + \delta_t + \varepsilon_{it} \tag{6}$$

In equation (5), the vector of parcel attributes and location characteristics $\boldsymbol{X}_{it}$ can be decomposed into five different categories: (1) the parcel-specific characteristics $\boldsymbol{L}_{it}$ such as soil quality, slope and elevation of the parcel; (2) the weather characteristics $\boldsymbol{W}_{it}$ such as temperature and precipitation; (3) the amenities and dis-amenities characteristics $\boldsymbol{A}_{it}$ such as proximities to a waterbody and Superfund site; (4) the urban influence characteristics $\boldsymbol{U}_{it}$ such as county-based population density, median household income, and proximity to cities; (5) agricultural market influence characteristics $\boldsymbol{M}_{it}$ such as cash rent and proximity to grain elevators, so that

$$\boldsymbol{X}_{it} = \boldsymbol{L}_{it} + \boldsymbol{W}_{it} + \boldsymbol{A}_{it} + \boldsymbol{U}_{it} + \boldsymbol{M}_{it} \tag{7}$$

Alternatively, we can define the following specification:

$$V_{it} = E_t \sum_{t=1}^{\infty} f(\boldsymbol{L}_{it}, \boldsymbol{W}_{it}, \boldsymbol{A}_{it}, \boldsymbol{U}_{it}, \boldsymbol{M}_{it}; r_t) \tag{8}$$

4.3.2 Empirical Model

The Hedonic price method, which was initially introduced by Griliches (1961) and further developed by Rosen (1974), has become the popular approach and has been widely employed in modeling the determinants of agricultural land values (Delbecq, et al., 2014, Dillard, et al., 2013, Zhang and Nickerson, 2015). The hedonic price method is known as a revealed preference method that the value of a parcel of agricultural land is a function of its attribute and characteristics. Numerous applications of hedonic models have been applied to examine the critical characteristics that affect farmland values. Guiling, et al. (2009), Delbecq, et al. (2014), and Zhang and Nickerson (2015) identify the extent of the urban fringe and its impact on agricultural land values. Furthermore, Bastian, et al. (2002), Wasson, et al. (2013), and Borchers, et al. (2014) investigate the effects of environmental amenities on agricultural land values. Other studies investigate the effect of other specific factors on the land values, including erosion control and drainage (Palmquist and Danielson, 1989), farmland preservation programs (Nickerson and Lynch, 2001), potential land development (Plantinga, et al., 2002), wildlife recreation income (Henderson and Moore, 2006), and wind facilities (Heintzelman and Tuttle, 2012).

Our empirical model with the hedonic price model is specified as a linear combination of parcel attributes and location characteristics that were defined previously. Under the hedonic price model, farmland is a differentiated product with a bundle of agricultural quality and location characteristics, and implicit prices can be estimated based on each characteristic. By substituting equation (8) into equation (7), we have

$$ln(V_{it}) = \beta_0 + \beta_L \boldsymbol{L}_{it} + \beta_W \boldsymbol{W}_{it} + \beta_A \boldsymbol{A}_{it} + \beta_U \boldsymbol{U}_{it} + \beta_M \boldsymbol{M}_{it} + \delta_t + \varepsilon_{it} \qquad (9)$$

The hedonic regression is formed with the log-linear specification. Because there is no clear theoretical guideline for the correct functional form for hedonic pricing models, a semi-log is preferred as a more flexible-form with unobserved attributes or presence of measurement error (Borchers, et al., 2014). The estimated regression estimated is defined as

$$\begin{aligned} &ln(land\ value_i) \\ &= \beta_1 Cash\ Rent_i + \beta_2 Acre_i + \beta_3 Acre_i^2 + \beta_4 Acre_i^3 + \beta_5 Clay_i \\ &+ \beta_6 Silt_i + \beta_7 Temp + \beta_8 Precipitation_i + \beta_9 Slope_i \\ &+ \beta_{10} Elevation_i + \beta_{11} Grain\ Elevator_i + \beta_{12} Superfund_i \\ &+ \beta_{13} City_i + \beta_{14} Waterbody_i + \beta_{15} Park_i + \beta_{16} Hospital_i \\ &+ \beta_{17} Population\ Density_i + \beta_{18} Median\ Income + \beta_{19} Resale_i \\ &+ \beta_{20} Immediate_i + \beta_{21} Building_i + \varepsilon_i \end{aligned} \qquad (10)$$

The dependent variable in this study is the per acre sales value.[43] Although the Box-Cox test could be applied to select the functional form, this study could not employ the Box-Cox test since some of the independent variables contain zero values. Furthermore, all independent variables are selected based on the Variance Inflation Factor (VIF) that tests multi-collinearity problems. Based on the VIF result, we find there are no severe multi-collinearity problems among the variables. In this study, we control the potential impacts

[43] In the ZTRAX, the sales price is entered in whole dollars and amounts $100 or less are ignored.

of outliers by considering observations between $150/acre and $15,000/acre. It is because prices that are too low (high) may indicate transactions among related individuals below (above) the market value (Guiling, et al., 2009).[44]

We begin with a specification that treats our sample as pooled cross-section data that assumes all sales transactions are independent. This is because we do not have enough repeated transactions, and as a result, any panel methods such as fixed or random effect estimation may not be adequate. In this regard, we take out the time subscription from equation (9) as shown in equation (10). However, we include a dummy variable to control how agricultural land values are associated with repeated sales compared to non-recurring transactions. Furthermore, we incorporate several dummy variables such as the month, year, and agricultural district to control for time-specific fixed effects and regional/locational heterogeneity, respectively. In addition, we estimate equation (10) by replacing the cash rent with NCCPI in order to compare the two models. One of the main objectives of this paper is to examine whether NCCPI could be a good indicator for the cash rent. If the model with NCCPI predicts agricultural land values relatively better than the model with cash rent, this result would support the idea that NCCPI might be a good indicator for cash rent.

4.4 Data

The primary source of data used in this study is the Zillow Transaction and Assessment Data (ZTRAX) provided by Zillow, an online real estate database company.

---

[44] The threshold used in this study is based on the previous studies. For example, Delbecq et al (2014) exclude observations when sales price is below $100/acre and above $20,000/ acre. In addition, Zhang and Nickerson (2015) limit observations if the estimated sales price for farmland value is between $1,000/acre and $20,000/acre. Likewise, the threshold to limit observation could be arbitrary. Our study tests with a different threshold that per acre sales price is between $100 and $1,000, and we find the results are robust.

The ZTRAX data includes information on approximately 374 million public records across more than 2,750 U.S. counties. Notably, the ZTRAX data covers more than 20 years of mortgages, foreclosures, auctions, and property taxes, especially for residential and commercial properties. In addition, all the corresponding property characteristics, geographical information, and prior valuation on approximately 200 million parcels are included across 3,100 counties in the assessor data. Utilizing the ZTRAX data, we geocode locations of the agricultural farmlands and identify the value of the farmland in Kentucky from 2009 to 2014.[45] In total, 3,845 transactions occurred for 3,546 parcels, indicating about 8% of the sample have sold more than once during the study period. Figure 4.3 shows the locations of agricultural land sold in Kentucky from 2009 to 2014, and it also shows the agricultural land value prices per acre that range from $2.42 to $901,315.80. As shown in figure 4.3, agricultural lands are widely and randomly distributed over the entire state.

The following other explanatory variables used in this study for parcel attributes and locations are discussed and explained with the previous studies on land value. Data on parcel attributes and location characteristics were collected mostly from the U.S. Department of Agriculture Natural Resources Conservation Services GeoSpatial Data Gateway (GeoSpatial Data Gateway 2018), including National Elevation Dataset, and Gridded Soil Survey Spatial Data (gSSURGO).

Soil quality is considered as an essential factor that influences farmland values in that farmland with higher soil quality leads to not only fewer production inputs and management time but also higher expected farming returns (Nickerson, et al., 2012). To

---

[45] We extract the agricultural land data from the ZTRAX based on the property land use classification.

control for soil quality, we obtain data on soil textures (e.g., percent clay, percent silt, and percent sand) from gSSURGO, which is provided by USDA National Resources Conservation Service (NRCS). The gSSURGO database provides greater spatial extents than the traditional SSURGO.[46] We include the National Elevation data (30-meter resolution) to calculate the elevation and slope. Some of the previous studies on land values include the slope such as (Borchers, et al., 2014, Zhang and Nickerson, 2015) and elevation such as (Buck, et al., 2014). The elevation could provide aesthetic qualities to an area; in addition, higher elevation could result in a shorter growing season with a high risk of crop damage from freezing (Vasquez, et al., 2002). For the slope, Palmquist and Danielson (1989) find that sales price for agricultural land is negatively related to the slope since steeper slop results in more erosion. A recent study by Borchers, et al. (2014) includes the slope as a measure of topography and finds a positive relationship between slope and farmland price especially for pastureland whereas the relationship is not statistically significant for the cropland. Moreover, Zhang and Nickerson (2015) find that the slope does not statistically influence the farmland values in western Ohio. In this regard, this study expects that the direction of the impact of slope and elevation on farmland values could be unknown.

The weather data such as precipitation and temperature were obtained from the Parameter-elevation Regressions on Independent Slopes Model (PRISM) climate Group.[47] The weather characteristics were included in that particular landscape, and climate features provide rich natural amenities (Borchers, et al., 2014, McGranahan, 1999). This study

[46] For more detail information about the gSSURGO, see https://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs142p2_052164.pdf
[47] http://prism.oregonstate.edu/

hypothesizes that precipitation and temperature will be significantly associated with agricultural land values. It is because crop profitability is heavily dependent on the weather condition in the growing season. If the weather condition is unstable from year-to-year, there could be high risk associated with the land for planting crops. For the weather variables, we calculate and use an average temperature and total precipitation from March to August in each year of the growing season.

This study includes urban influence factors because previous research such as Huang, et al. (2006), Livanis, et al. (2006), Zhang and Nickerson (2015), and Burns, et al. (2018) suggest that farmland values are positively associated with near urban and developed areas. This positive association could be explained by the fact that the farmlands near urban lead to higher return by reallocating production from commodity-oriented agriculture to higher-valued commodities (Livanis, et al., 2006). In addition, farmland values could be positively influenced by increased access to markets and customers but also proximity to population centers (Nickerson, et al., 2012). In this regard, we include the proximity to the cities in Kentucky by calculating the minimum distance to the closest city. All the locations of the major city are obtained from the Environmental System Research Institute (ESRI 2018). In addition to the urban influence, we incorporate the locations of the grain elevators. We obtain the geographical locations of the grain markets from the Department of Agricultural Economics at the University of Kentucky.[48] We then measure the proximity to the closest locations of the grain elevator. This study hypothesizes the positive relationship between the land values and the locations of the grain elevators

[48] The map of Kentucky Grain Markets is originally generated by Dr. Jordan M. Shockley who is an Assistant Extension Professor in the Department of Agricultural Economics at the University of Kentucky.

since lands located closer to grain elevators tend to be valued more highly (Burns, et al., 2018, Nickerson, et al., 2012, Zhang and Nickerson, 2015).

The existing literature on farmland values show that recreation and natural amenities are positively associated with the farmland values (Bastian, et al., 2002, Borchers, et al., 2014, Nickerson, et al., 2012, Wasson, et al., 2013). Based on the previous studies, this study includes the locations of the waterbody, recreational park, hospital, and Superfund sites.[49] The water body boundary, locations of the recreational park and hospital are obtained from ESRI, which provides many data layers for U.S. Census, government and non-government, and commercial geographies (Borchers, et al., 2014).[50] Based on ESRI-provided landmark and recreation dataset, we measure the nearest distances to these features. For the Superfund site, we collect data such as longitude and latitude of each Superfund site in Kentucky from the U.S. Environmental Protection Agency (EPA) and calculate the distance to the closest Superfund site. The Superfund sites, however, are considered as a dis-amenity because it represents the lands or areas that are contaminated and hazardous with toxic wastes. Previous studies such as Fischhoff (2001), Davis (2004), Messer, et al. (2006), and Gamper-Rabindran and Timmins (2013) provide evidence of an inverse relationship between the property values and Superfund sites. Therefore, we hypothesize that agricultural land values are negatively affected by Superfund sites.

Other control variables used in this study include median household income and population density. Agricultural lands located in higher income and high population density county might reveal greater economic opportunities for residents (Borchers, et al.,

[49] We considered other variables such as the locations of nearest golf course and college/university. However, we excluded those variables due to the multicollinearity problem.
[50] ESRI is also known as the supplier of the Geographic Information System software ArcGIS.

2014). In this study, we use county-level measures of median household income and total population obtained from the USDA Economic Research Service (USDA, ERS).[51] Huang, et al. (2006) and Borchers, et al. (2014) find that farmland values in both cropland and pasture are positively influenced by median household income and population intensity index. Furthermore, Ifft, et al. (2015) and Zhang and Nickerson (2015) show there is a positive relationship between farmland values and total population. This study, therefore, hypothesizes that farmland values in Kentucky are positively associated with county-level measures: median household income and total population density.

For an observed measure of agricultural return, this study includes the cash rent as an explanatory variable. The cash rent used here is the aggregated county-level data and is obtained from the U.S. Department of Agriculture, National Agricultural Statistics Service (USDA, NASS). This study only employs and focuses on the non-irrigated cash rent rather than irrigated rent since agricultural lands in Kentucky are mostly non-irrigated.[52] As we discussed in the introduction, one of the main objectives of this paper is to test whether NCCPI could be a good indicator for the cash rent. In this regard, we utilize the National Commodity Crop Productivity Index (NCCPI). NCCPI is a national soil interpretation and generated in the National Soil Information System (NASIS). It relates to the ability of soils, landscape, and climates to enhance crop productivity (Dobos, et al., 2008).[53] Only a few recent studies such as Delbecq, et al. (2014) and Zhang and Nickerson (2015) include the

---

[51] The county population density is calculated by dividing the county total population by county square miles.

[52] The cash rent data is only available at the county level from 2009 to 2014. Since the cash rent is the one of the main variables of interest, we limit our sample up to 2014. The cash rent for irrigated, we find there is almost no information.

[53] More detail information on NCCPI is available at https://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/16/nrcs143_020559.pdf

NCCPI to examine the farmland values in that NCCPI provides a measure of potential returns from the production of agricultural goods and services. Their findings demonstrate that farmland values are positively and significantly affected by NCCPI. In this study, we obtained NCCPI through use of gSSURGO. We extracted the values of NCCPI from the raster (see figure 4.2) and spatially joined to each parcel location using ArcGIS. The values of NCCPI indices are numbers ranging from 0 (least productive) to 1 (most productive).

Finally, this study incorporates different land classifications. In the ZTRAX, agricultural land is classified with 7 different property land uses: general agriculture, farm non-irrigated or dry, timberland/forestry/trees, livestock, rural improves (non-residential), miscellaneous structures, and unimproved vacant lands. We combine general agriculture, farm, and livestock as an immediate land and combine the rest as non-immediate land.[54] This study hypothesizes that land values in Kentucky could be positively related to the immediate lands compared to the non-immediate lands.

4.5 Results and Discussions

Table 4.1 shows the descriptive summary statistics for the dependent and covariates. The average parcel size is 46.77 acres; it has a price per acre of $3,599.93, and is predominantly classified as immediate lands (74%). Agricultural lands in Kentucky, on average, are composed of about 59% silt, 25% clay, and 16% sand. Only 7% of the sales transactions are made more than once from 2009 to 2014. On average, total precipitation and mean temperature from March to August are 29.50mm and 66.15 °F, respectively.

[54] The immediate lands represent the lands that could be converted or transferred to generate profits, whereas the non-immediate lands that could not be immediately converted for the profit generation.

Furthermore, the average county-level cash rent is $79.19, and average commodity crop productivity of parcels based on NCCPI are particularly below average (0.44).

Table 4.2 presents our main regression results. The regressions correspond to equation (10) and are reported with the model (1) with cash rent and model (2) with NCCPI. All reported models include agricultural district dummies, year dummies, and month dummies. Additionally, all models report robust standard errors clustered at the county level. Because the log-linear specification was used, the reported coefficients could be interpreted as the percentage change in per acre land value with a one-unit changes in the explanatory variable.

Based on model (1) with cash rent (column 2 in Table 4.2), the results suggest that farmland values in Kentucky are positively associated with cash rent, proximity to a grain elevator, county median income, county population density, and immediate lands. On the other hands, the land values are negatively associated with parcel size, repeated sales, slope, and proximity to the Superfund site. The model suggests that county-level cash rent has a statistically significant impact on land values, with an approximately 0.2% increase in land values with a $1/acre increase in cash rent. This might be because cash rents generally explain the amount paid per acre based on the measure of the productivity of the land. If the parcels are sold more than once in our study period, from 2009 to 2014, the values of the land are negatively associated with the repeated transactions compared to the single transaction. This negative effect might be explained that if agricultural lands are sold in the market more than once for a short time period, the lands are treated as vacant lands (i.e., no operation or management). This implies that the productivity of lands is likely to be lower and as a result lower the land values. Parcels one mile close to the gain elevator have a 0.4% increase in land values and this difference is significant. This positive relationship can be explained factors such as lower transportation cost and better inventory management. This finding is also supported by a study of Nickerson, et al. (2012). Results from the variable of immediate suggest that lands with general agriculture, farmland, and livestock are associated with 13.4% higher land values compared to lands with non-residential, miscellaneous structures, and unimproved vacant lands. This difference is also significant.

The estimates for parcel size (acre) show that a one-acre increase in parcel size is associated with a 1.2% decrease in land values. With the different functional forms, such as acre^2 and acre^3, this study finds there is a non-linear relationship between the land value and parcel size. Land values begin with a decreasing relationship with size, but it increases and then decreases again as the parcel size increases. Figure 4.4 shows the marginal effect of acres. We calculate the threshold points where the effect of acre changes from negative to positive and vice versa based on the following equation: $\hat{y} = 3.803646 - 0.0117155x + 0.0000359x^2 - 0.0000000285x^3$ where $\hat{y}$ = predicted land values and x = acre. By taking the partial derivative $\hat{y}$ respect to acre, $\frac{d\hat{y}}{dAcre} = -0.0117155 + 0.0000718x - 0.0000000855x^2 = 0$, we find that the threshold points are 221.70 and 618.07. With the threshold points, agricultural land values in Kentucky decrease until parcel size reaches 221.70 acres and continues to increase until 618.07 acres. It then starts to decrease. At stage 1 where the parcel size is below 221.70 acres, agricultural lands could be considered as hobby farms or hunting farms; in other words, people buy or lease lands for other purposes rather profits on lands. At stage 2 where the parcel size is between 221.70 and 618.07, the positive impact of parcel size on land values could be explained by increasing returns to scale or efficient land operation. At stage 3 where the parcel size is above 618.07, the negative relationship between land values and parcel size could be due to decreasing returns to scale or capital constraint. Since parcel size is relatively large in this stage, people might not want to buy or rent the lands; in addition, the production process in large lands may not be inefficient.

Several locational characteristics such as distance to the nearest city and recreational park have no significant impact on land values except for the proximity to the closest hospital: one mile closer to a hospital is associated with a 1.1% increase in land values. The insignificant effects of the locational characteristics can be explained by potential correlations between variables even though all locational characteristics are included in the model based on the VIF test. These results are similar to a previous study of Borchers, et al. (2014) in that they find the distance to the recreational waterbody and nearest park do not influence land values. Interestingly, we find that agricultural land values are negatively associated with the proximity to the closest Superfund sites: one mile close to a Superfund site results in a 0.3% decrease in land values. The impact of Superfund site on local property values, especially housing values, have been extensively investigated, and the vast of previous studies show that property values are negatively influenced by proximity to Superfund sites (Boyle and Kiel, 2001, Farber, 1998, Kiel and Williams, 2007). This finding will contribute to the existing literature on agricultural land values in that the proximity to the nearest Superfund site could be one of the critical determinants of land values. This study also finds that a one-degree increase in the slope of parcel results in a 1.8% decrease in land values. Lands with steep slope can lead to excessive erosion without improved production practices (which can be costly). Various county-level characteristics such as median household income and population density are associated with higher agricultural land values.

As we discussed before, we estimate equation (10) by replacing cash rent with NCCPI in order to compare which model predicts better. The model (2) in table 4.2 shows the regression results with NCCPI. Compared to the results of the regression model with the cash rent, the results with NCCPI are qualitatively similar to the regression model with cash rent although the coefficient estimates of interest vary slightly in terms of magnitude. The positive and significant effect of NCCPI on land values in this study is consistent with Zhang and Nickerson (2015). The reported R^2 in table 4.2 for model 1 with cash rent and model 2 with NCCPI is 33.6 and 33.7, respectively. Although R^2 could suggest which model predicts better than the other, we conduct several model fit tests, and test results are reported in table 4.3. As shown in table 4.3, we find that the regression model with NCCPI predicts relatively better than the model with cash-rent. In particular, the rule of thumb to find the better model is based on smaller values of information criteria (IC) and higher values of R^2.

This study also conducts a validation test of how accurately one model predicts land values relative to the other. For the validation test, we employ the following steps. First, we select 10% of the data using a random sampling process, called a hold-out sample. Second, we estimate the regression model with the remaining dataset (90% of data, called the training sample). Third, we predict using the hold-out sample then calculate the residuals in percentage term by taking the difference between actual values and predicted values. Fourth, we compare the residuals between the two models at the top 95%. By employing the out-of-sample validation, this study finds that 95% of the estimates from the regression models with cash rent and NCCPI are within 34.194% and 32.178% of the true values, respectively. For robustness, we resize the test sample size by 20% and 30%, and we find that land values with NCCPI are predicted 0.61% and 0.13% better than cash rent, respectively. This result provides suggestive evidence that agricultural land values are predicted relatively better with NCCPI rather than cash rent. The better prediction with NCCPI might be due to the fact that NCCPI is disaggregated data at the parcel level, whereas cash rent is aggregated at the county-level.[55] Although cash rent could be a key determinant to explain land values in that it has more variations over time, our study shows NCCPI should be considered a good indicator of land values. Currently, available NCCPI data does not provide any variations over time. This is because inherent productivity is considered almost invariant over time (Dobos, et al., 2008).

4.6 Conclusions

This study provides the first empirical examination of land values using the individual transaction data, Zillow Transaction and Assessment Data (ZTRAX) provided by Zillow. The main focus of the research is on three important research questions namely i) Are determinants of land values in Kentucky, where agricultural lands are heterogeneity, consistent with the previous studies on agricultural or farmland values? ii) What are the direction and size of the effect of parcel size on agricultural land values? iii) Could NCCPI be considered an acceptable indicator for agricultural land values compared to cash rent? The analysis is based on pooled OLS regression under the hedonic price model.

[55] Especially for Kentucky, the county-level cash rent data for 2015 is not publicly available.

Findings of this study provide evidence that factors influencing agricultural land values in states with heterogeneous agricultural lands such as Kentucky are not different from other states with relatively homogeneous agricultural lands. In contrast to the previous studies, several findings in this study may have important implications and therefore contribute to the growing existing literature of land values. We particularly find that the agricultural land values decrease by 19.2% if lands are sold more than once compared to the single transaction. This suggests that land management practice might be needed to sustain the land quality and productivity when lands are in the market. When agricultural lands are categorized with more specific land classifications, we find that lands with general agriculture, farmland, and livestock are associated with 8.4% increase in land values. Furthermore, if there is a building on the parcel, the agricultural land values are increased by 30.3% compared to the land without the building. Interestingly, we find that agricultural land values are negatively associated with the proximity to the closest Superfund sites. To the best of our knowledge, no studies investigate the relationship between Superfund site and land values even though a vast literature has examined the impact of Superfund site on residential property values.

Our results also provide suggestive evidence that there is a non-linear relationship between parcel size and land values. We specifically find two threshold points where the marginal effect of parcel size varies from negative to positive and vice versa. By knowing the non-linear relationship and threshold points, it may provide important baseline information for land owners to manage the efficient land allocation in order to generate more revenues: For instance, there is a negative relationship between parcel size and farmland value if parcel size is larger than 618.07 acres. This suggests that land management or allocation from larger parcels to smaller parcels could generate higher revenue. Furthermore, we find that a disaggregated NCCPI at the parcel-level is an acceptable indicator to estimate agricultural values compared to an aggregated cash rent at the county-level. Although both cash rent and NCCPI show a significantly positive impact on land values with similar statistical power, NCCPI is found to predict land values better than county-level cash rent. This finding could imply that using aggregated data on cash rent may be substitutable with the disaggregated data, NCCPI, not only to investigate the individual transaction data at the field-level but also to mitigate any potential aggregation bias problem. It also contributes to the existing literature on analyzing agricultural land values under the hedonic model in that price information such as cash rent should not be necessary to be included.

This study has several limitations. First, our main dataset of ZTRAX is a rich dataset on all individual transactions for both property and land values at the parcel. Nevertheless, researchers who use the ZTRAX data should make sure that the observations and information obtained from the ZTRAX sufficiently cover and represent the study areas since we find some states unreasonably have lack of observations. Second, we find that NCCPI could be a good indicator of agricultural land values since the estimated regression model with NCCPI predicts relatively better than the model with cash rent. However, this result may not be consistent with the field-level cash rents based on the availability of data and may vary in the different states due to the large variations in cash rents within the state.

4.7 Tables and Figures

Table 4.1 Descriptive Summary Statistics (N=3,266)

Variable	Variable Description/Definition	Mean	Std. Dev.	Min	Max
Log Price	Agricultural land sales per acre (in log)	7.77	0.98	5.01	9.62
acre	Parcel size in acre	46.55	58.67	0.18	872
acre2	Parcel size in acre square	5608.82	29379.19	0.03	760384
acre3	Parcel size in acre cubic	1684381	21300000	0	663000000
Cash Rent	County-level Non-Irrigated Cash Rent (dollar)	79.19	36.81	24.00	210
NCCPI	National Commodity Crop Productivity Index	0.44	0.25	0.01	0.94
Repeated Sales	= 1 if parcel sold more than once, 0 otherwise	0.07	0.26	0	1
Clay	Soil with a combination of Clay in parcel (percent)	24.90	9.01	4.00	58.00
Silt	Soil with a combination of Silt in parcel (percent)	58.81	13.30	15.70	82.00
Precipitation	Total precipitation from March to August (mm)	29.50	6.38	11.05	45.61
Temperature	Mean temperature from March to August (°F)	66.15	2.50	60.97	72.10
Slope	Slope of Parcel (degree)	6.08	5.61	0.00	41.97
Elevation	Slope of Elevation (m)	226.52	67.40	96.40	609.72
Grain Elevator	Distance to nearest gain elevator (miles)	18.98	13.85	0.08	91.12
Superfund	Distance to nearest Superfund site (miles)	32.00	19.38	0.15	84.84
City	Distance to nearest city (miles)	17.89	12.74	0.78	81.86
Waterbody	Distance to nearest waterbody for recreation (miles)	2.83	2.35	0.01	15.72
Park	Distance to nearest national, state, or local park (miles)	5.29	3.06	0.07	19.78
Hospital	Distance to nearest hospital (miles)	9.24	4.18	0.13	22.11
Median Income	County-level median household income (in thousand dollar)	41.21	8.99	23.16	80.87
Population Density	County-level population density (percent)	164.31	229.44	22.05	1913.88
Immediate	= 1 if parcel is agriculture, farm, or livestock, 0 otherwise	0.74	0.44	0	1
Building	= 1 if there is a building located on parcel	0.21	0.41	0	1

Table 4.2 Regression Results

Variables	**Model 1** (Cash Rent)	**Model 2** (NCCPI)
Acre	-0.012***	-0.012***
	(0.001)	(0.001)
Acre2	0.00004***	0.00004***
	(0.00005)	(0.00005)
Acre3	-0.000***	-0.000***
	(0.000)	(0.000)
Cash Rent	0.002**	—
	(0.001)	—
NCCPI	—	0.264***
	—	(0.081)
Repeat	-0.192**	-0.197**
	(0.078)	(0.077)
Clay	-0.002	0.001
	(0.002)	(0.002)
Silt	0.001	0.0004
	(0.001)	(0.002)
Precipitation	-0.001	-0.001
	(0.005)	(0.005)
Temperature	0.051	0.056
	(0.027)	(0.027)
Slope	-0.018***	-0.015***
	(0.004)	(0.004)
Elevation	0.0004	0.0006
	(0.0005)	(0.0005)
Grain Elevator	-0.004*	-0.004*
	(0.002)	(0.002)
Superfund	0.003**	0.002*
	(0.001)	(0.001)
Median Income	0.022***	0.023***
	(0.004)	(0.004)
Population Density	0.0004***	0.0004***
	(0.00006)	(0.00006)

Note: Robust standard errors in parentheses. *** $p<0.001$, ** $p<0.01$, * $p<0.05$

Table 4.2 (Continued)

Variables	**Model 1** (Cash Rent)	**Model 2** (NCCPI)
City	-0.003	-0.003
	(0.002)	(0.002)
Waterbody	-0.011	-0.011
	(0.008)	(0.009)
Parks	-0.009	-0.008
	(0.009)	(0.009)
Hospital	-0.011**	-0.011*
	(0.005)	(0.005)
Immediate	0.085*	0.077
	(0.047)	(0.047)
Building	0.303***	0.301***
	(0.057)	(0.057)
Constant	3.804**	3.355*
	(1.862)	(1.904)
Observations	3,266	3,266
R-squared	0.336	0.337
District FE	Yes	Yes
Year FE	Yes	Yes
Month FE	Yes	Yes

Note: Robust standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05

Table 4.3 Comparison Measure of Fit between Two Models

	Model 1 (Cash Rent)	**Model 2** (NCCPI)
Log-likelihood		
Model	-3896.765	-3893.975
Intercept-only	-4565.198	-4565.198
Chi-square		
Deviance (df=3324)	7793.530	7787.949
R^2		
R^2	0.336	0.337
Adjusted R^2	0.327	0.328
McFadden	0.146	0.147
McFadden (adjusted)	0.137	0.138
Cox-Snell/ML	0.336	0.337
Cragg-Uhler/Nagelkerke	0.358	0.359
IC		
AIC	7877.530	7873.949
AIC divided by N	2.412	2.411
BIC (df=42)	8133.365	8135.876

Notes: The value of Cox-Snell/ML represents the R-Squared that is calculated by $R^2 = 1 - \left\{\frac{L(M_{Intercept})}{L(M_{Full})}\right\}^{2/N}$ where $L(M_{Intercept})$ is the log likelihood of the intercept model and $L(M_{Intercept})$ is the log likelihood of the full model. The value of Cragg-Uhler/Nagelkerke also represents the R-Squared, which is calculated by $R^2 = 1 - \frac{\left\{\frac{L(M_{Intercept})}{L(M_{Full})}\right\}^{2/N}}{1-L(M_{Intercept})^{2/N}}$. AIC and BIC represent Akaike's Information Criterion and Bayesian Information Criterion, respectively. To compare between two models (Model 1 and Model 2), the smaller the deviance, the better the model fit. For the R^2, the higher the R^2, the better the model fit. Finally, the smaller AIC and BIC show better model fit.

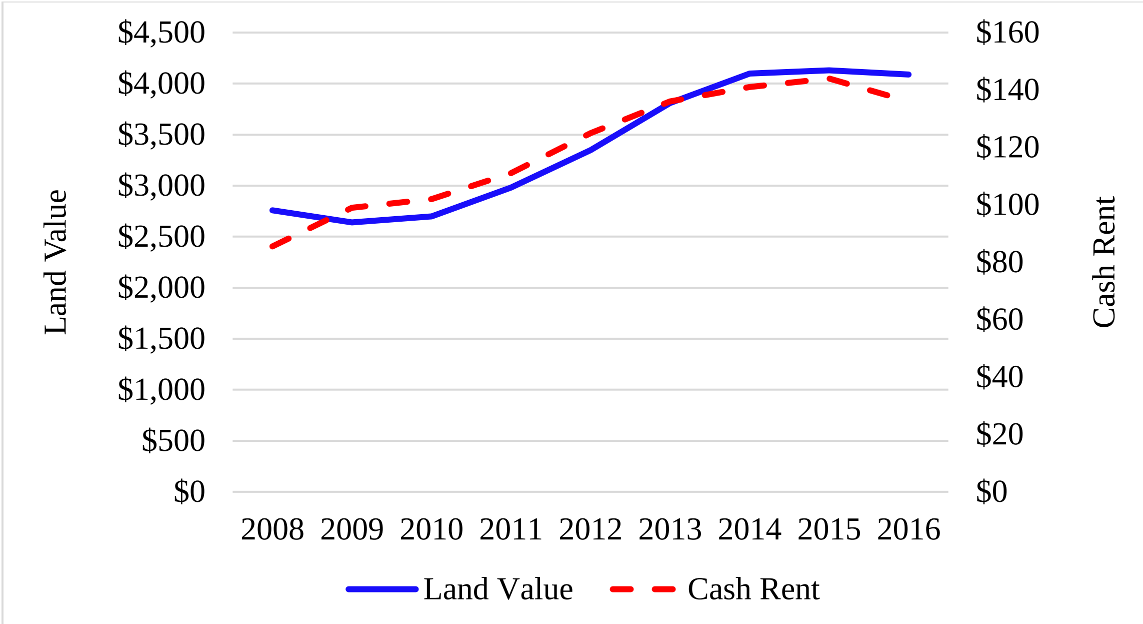


Figure 4.1 Total Value of U.S. Hemp Imports, 2010-2015

Source: USDA NASS

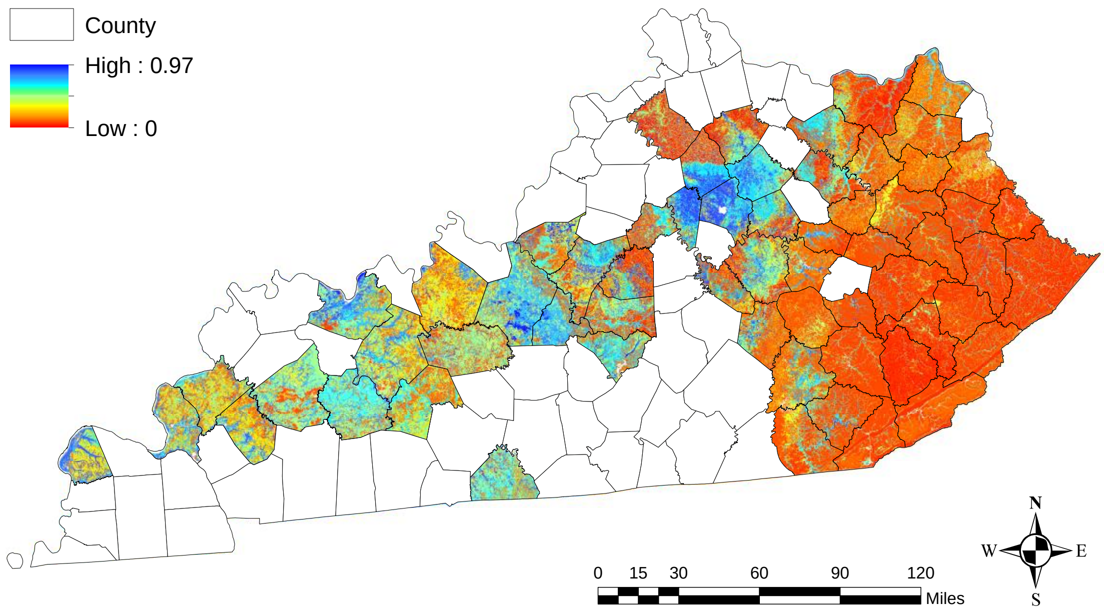


Figure 4.2 National Commodity Crop Productivity in Kentucky (in %)

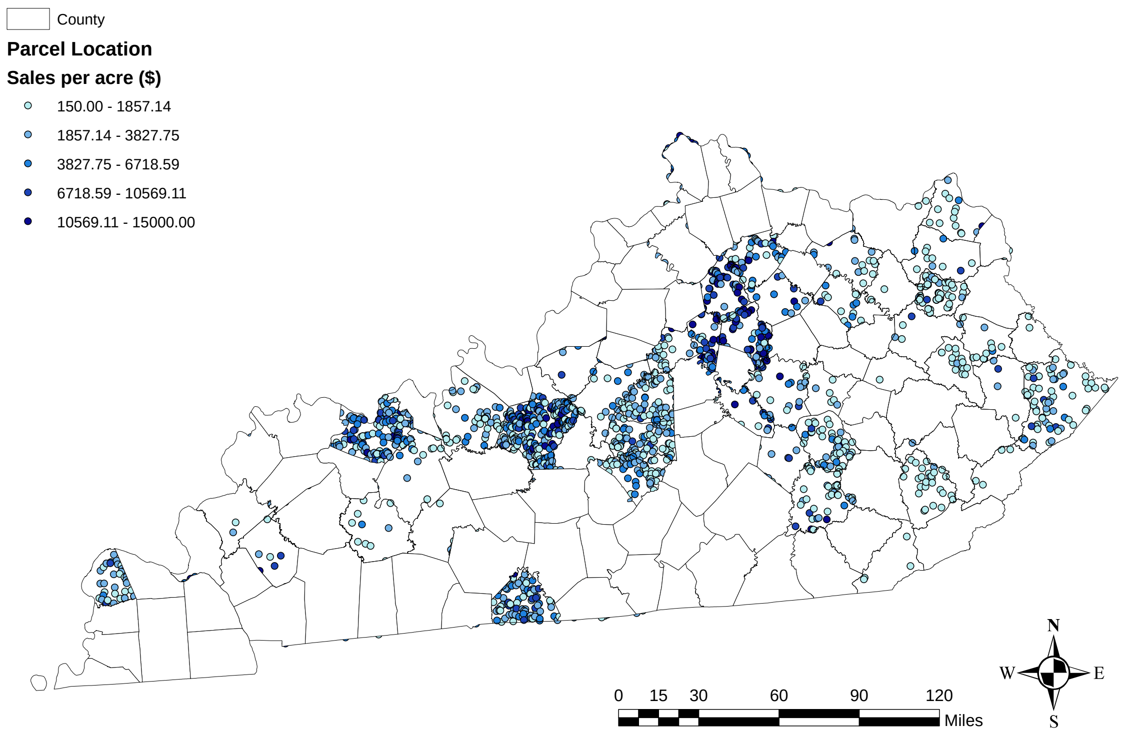


Figure 4.3 Location of Parcel Sold in Kentucky from 2009 to 2014

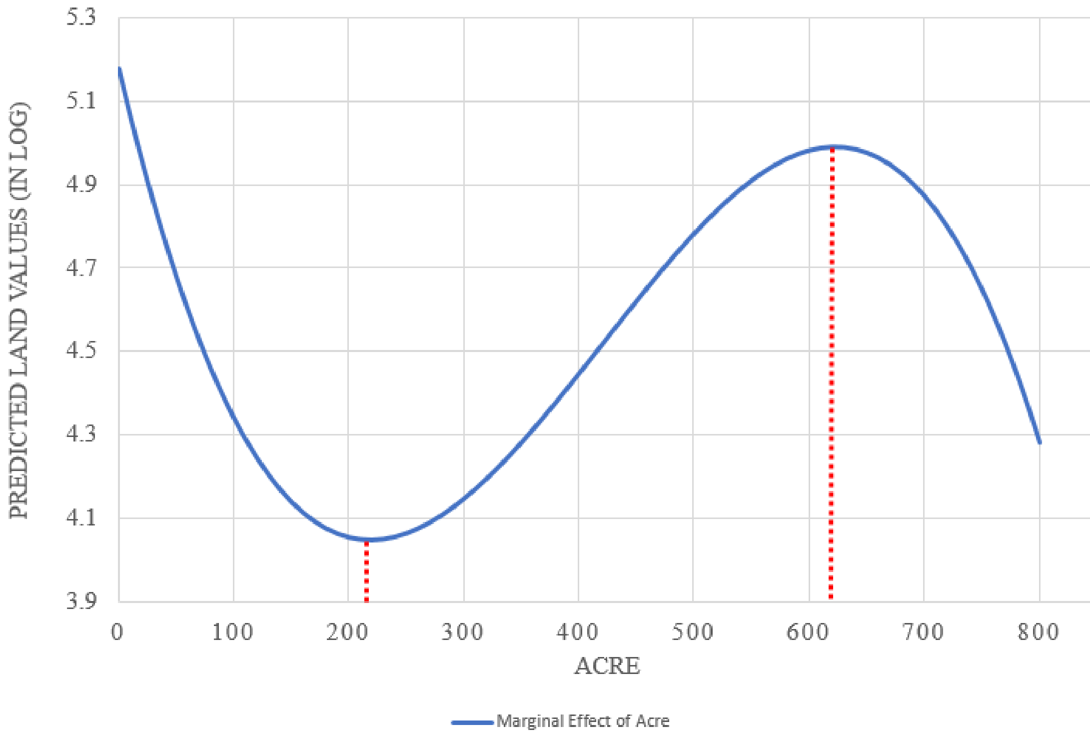


Figure 4.4 Marginal Effect of Acre

CHAPTER 5. SUMMARY AND CONCLUSIONS

This dissertation combines large scale datasets to evaluate crop prediction, land values, and consumption of crop being considered to advance a sustainable bioeconomy. Chapter 2 proposes a novel application of the multinomial logit (MNL) model to estimate the conditional transition probabilities of crop choice and forecast distribution of total acreages by crop type for the state of Kentucky from 2010 to 2015. The Cropland Data Layer (CDL) data is primarily utilized and merged with the Common Land Unit (CLU) dataset. Findings show that corn is more likely to be followed by soybeans, whereas monoculture crops such as tobacco and alfalfa are more likely to be planted in consecutive years. In addition, the forecasted distributions based on the simulation exercise show wider distributions for corn and soybeans, whereas narrower distributions for tobacco, wheat, and alfalfa. The wide distribution in corn acreage indicates a high likelihood of above average nutrient run-off since, on average, corn receives nitrogen and phosphorous applications. In addition, the tighter distributions in alfalfa and tobacco acreages can be explained that alfalfa is a perennial crop and tobacco is contacted crop. The forecasted distributions can be used and applied in various fields of research and will contribute to policy analysis. For instance, the distribution can be used to make probability statements related to the ability of producers to incorporate new crops such as hemp into the land-use rotations as well as using distributions of land-use to generate distributions of soil erosion, nitrogen run-off and other soil and water quality indicators.

Chapter 3 investigates the critical sociodemographic factors that are associated with increasing hemp consumption and examines their effect on total expenditure in the U.S. by utilizing Nielsen's consumer panel data from 2008 to 2015. We find that sociodemographic characteristics, especially income, age, and education, play an important role in purchasing and explaining the demand for different categories of hemp products. Specifically, households with higher income and education are positively associated with the probability and consumption level of hemp products. In addition, households with higher education are more likely to buy hemp products, and of those households are significantly associated with a higher level of consumption except hemp nuts. However, the role of the age of household head shows mixed results with respect to consumption decisions and consumption levels across the products. These findings provide insights into a more targeted marketing strategy for hemp industries to attract new consumers and increase sales to current consumers. Furthermore, findings in this study fill the knowledge gap on a new agricultural crop that is increasing consumption and production in the U.S. Since there are many unknowns about everything from hemp production to its marketing channels, a basic understanding of consumer profiles will provide a starting point for these discussions.

Chapter 4 utilizes a new rich dataset, Zillow Transaction and Assessment Data (ZTRAX) provided by Zillow, to investigate the factors that affect agricultural land values in Kentucky from 2009 to 2014. Agricultural lands in Kentucky are relatively heterogeneous compared to other states like Iowa, Illinois, and Nebraska, where agricultural lands are homogeneous in that those states are known as corn-belt states. Findings show that factors influencing land values in Kentucky are not different from other states discussed in the previous and existing literature. However, this study finds that agricultural land values decrease by 19.3% if lands are sold more than once compared to a single transaction, suggesting land management practice might be needed to sustain the land quality and productivity when lands are in the market. We also find that land values are positively associated with specific land classifications: general agriculture, farmlands, and livestock. Furthermore, our results provide suggestive evidence that there is a non-linear relationship between parcel size and land values: agricultural land values in Kentucky decrease until parcel size reaches 218.64 acres and continue to increase until 622.25 acres. It then starts to decrease. Knowing the non-linear relationship and specific threshold points might provide important policy-oriented implication to manage the efficient land allocation and improve agricultural land values by reducing land use conflict resolutions. Finally, we find that a disaggregated NCCPI at the parcel-level is an acceptable indicator to estimate agricultural values compared to an aggregated cash rent at the county-level. This implies that price information such as cash rent should not be necessary to be included in the hedonic price model for analyzing the agricultural values.

REFERENCES

Adjemian, M.K., and A. Smith. 2012. "Using USDA forecasts to estimate the price flexibility of demand for agricultural commodities." *American Journal of Agricultural Economics* 94:978-995

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.

Aurbacher, J., and S. Dabbert. 2011. "Generating crop sequences in land-use models using maximum entropy and Markov chains." *Agricultural Systems* 104:470-479.

Bastian, C.T., et al. 2002. "Environmental amenities and agricultural land values: a hedonic model using geographic information systems data." *Ecological Economics* 40:337-349.

Baker, W.L. 1989. "A review of models of landscape change." *Landscape ecology* 2:111-133.

Barnard, C.H. 2000. "Urbanization affects a large share of farmland." *Rural Conditions and Trends* 10.2: 57-63.

Bell, E.J. 1974. "Markov analysis of land use change—an application of stochastic processes to remotely sensed data." *Socio-Economic Planning Sciences* 8:311-316.

Blaug, M. 1997. *Economic theory in retrospect*: Cambridge university press.

Boryan, C., et al. 2011. "Monitoring US agriculture: the US department of agriculture, national agricultural statistics service, cropland data layer program." *Geocarto International* 26:341-358.

Borchers, A., J. Ifft, and T. Kuethe. 2014. "Linking the price of agricultural land to use values and amenities." *American Journal of Agricultural Economics* 96:1307-1320.

Brown, D.G., B.C. Pijanowski, and J. Duh. 2000. "Modeling the relationships between land use and land cover on private lands in the Upper Midwest, USA." *Journal of Environmental Management* 59:247-263.

Buck, S., M. Auffhammer, and D. Sunding. 2014. "Land Markets and the Value of Water: Hedonic analysis using repeat sales of farmland." *American Journal of Agricultural Economics* 96:953-969.

Burgess, M.H., P.R. Miller, and C.A. Jones. 2012. "Pulse crops improve energy intensity and productivity of cereal production in Montana, USA." *Journal of sustainable agriculture* 36:699-718.

Burns, C., et al. "Farmland Values, Land Ownership, and Returns to Farmland, 2000-2016."

Capps, O., et al. 1999. "Examining packer choice of slaughter cattle procurement and pricing methods." *Agricultural and Resource Economics Review* 28:11-25.

Carrión-Flores, C.E., A. Flores-Lagunes, and L. Guci (2009) "Land use change: a spatial multinomial choice analysis." In *Agricultural and Applied Economics Association 2009 Annual Meeting.*

Castellazzi, M., et al. 2008. "A systematic representation of crop rotations." *Agricultural Systems* 97:26-33.

Cherney, J.H., and E. Small. 2016. "Industrial hemp in North America: Production, politics and potential." *Agronomy* 6:58.
Croissant, Y. 2012. "Estimation of multinomial logit models in R: The mlogit Packages." *R package version 0.2-2. URL: http://cran. r-project. org/web/packages/mlogit/vignettes/mlogit. pdf*.
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.
David, R. 1817. "On the principles of political economy and taxation." *publicado en*.
Davis, L.W. 2004. "The effect of health risk on housing values: Evidence from a cancer cluster." *American Economic Review*:1693-1704.
de Barros Dias, F. 2017. "US yield forecasting using crop condition ratings."
Delbecq, B.A., T.H. Kuethe, and A.M. Borchers. 2014. "Identifying the extent of the urban fringe and its impact on agricultural land values." *Land Economics* 90:587-600.
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.
Dillard, J.G., et al. 2013. "The Impacts of the Tax-Deferred Exchange Provision on Farm Real Estate Values." *Land Economics* 89:479-489.
Dobos, R., H. Sinclair, and K. Hipple. 2008. "User guide national commodity crop productivity index (NCCPI) version 1.0." *US Department of Agriculture, Natural Resources Conservation Service*.
Domenich, T., and D. McFadden. 1975. "Urban travel demand: a behavioural approach." *Worth-Holland, Amsterdam*.
Ehrensing, D.T. 1998 "Feasibility of industrial hemp production in the United States Pacific Northwest." Corvallis, Or.: Agricultural Experiment Station, Oregon State University.
Egelkraut, T.M., et al. 2003. "An evaluation of crop forecast accuracy for corn and soybeans: USDA and private information agencies." *Journal of Agricultural and Applied Economics* 35:79-95.
Farber, S. 1998. "Undesirable facilities and property values: a summary of empirical studies." *Ecological Economics* 24:1-14.
Featherstone, A.M., and T.G. Baker. 1987. "An examination of farm sector real asset dynamics: 1910–85." *American Journal of Agricultural Economics* 69:532-546.
Fike, J. 2016. "Industrial Hemp: Renewed Opportunities for an Ancient Crop." *Critical Reviews in Plant Sciences* 35:406-424.
Fischhoff, B. 2001. "Defining stigma." *Risk, Media and Stigma–Understanding Public Challenges to Modern Science and Technology*:361-368.
Flanders, A., F.C. White, and C.L. Escalante. 2004. "Equilibrium of Land Values from Agricultural and General Economic Factors for Cropland and Pasture Capitalization in Georgia." *Journal of Agribusiness* 22:49.
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."

Gamper-Rabindran, S., and C. Timmins. 2013. "Does cleanup of hazardous waste sites raise housing values? Evidence of spatially localized benefits." *Journal of environmental economics and management* 65:345-360.

Golden, J.S., et al. 2015. "An economic impact analysis of the US biobased products industry: A report to the Congress of the United States of America." *Industrial Biotechnology* 11:201-209.

Good, D., and S. Irwin. 2015. "Progression of USDA Corn and Soybean Acreage Estimates and Prospects for Final Estimates for 2015." *farmdoc daily*.

---. 2011. "USDA corn and soybean acreage estimates and yield forecasts: dispelling myths and misunderstandings." *Marketing and Outlook Brief* 2:11-02.

Greene, W.H. 2003. "Econometric analysis, 5th." *Ed.. Upper Saddle River, NJ*.

Griliches, Z. (1961) "Hedonic price indexes for automobiles: An econometric of quality change." In *The Price Statistics of the Federal Goverment.* NBER, pp. 173-196.

Guiling, P., B.W. Brorsen, and D. Doye. 2009. "Effect of urban proximity on agricultural land values." *Land Economics* 85:252-264.

Halvorson, A.D., S.J. Del Grosso, and C.A. Reule. 2008. "Nitrogen, tillage, and crop rotation effects on nitrous oxide emissions from irrigated cropping systems." *Journal of Environmental Quality* 37:1337-1344.

Hardie, I.W., and P.J. Parks. 1997. "Land use with heterogeneous land quality: an application of an area base model." *American Journal of Agricultural Economics* 79:299-310.

Hayes, M., and W. Decker. 1996. "Using NOAA AVHRR data to estimate maize production in the United States Corn Belt." *Remote Sensing* 17:3189-3200.

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

Heintzelman, M.D., and C.M. Tuttle. 2012. "Values in the wind: A hedonic analysis of wind power facilities." *Land Economics* 88:571-588.

Henderson, J., and S. Moore. 2006. "The capitalization of wildlife recreation income into farmland values." *Journal of Agricultural and Applied Economics* 38:597-610.

Hendricks, N.P., et al. 2014a. "The environmental effects of crop price increases: Nitrogen losses in the US Corn Belt." *Journal of environmental economics and management* 68:507-526.

Hendricks, N.P., A. Smith, and D.A. Sumner. 2014b. "Crop supply dynamics and the illusion of partial adjustment." *American Journal of Agricultural Economics* 96:1469-1491.

Huang, H., et al. 2006. "Factors influencing Illinois farmland values." *American Journal of Agricultural Economics* 88:458-470.

Ibendahl, G., and T. Griffin. 2013. "The Connection Between Cash Rents and Land Values." *Journal of ASFMRA*:239-247.

Ifft, J., T. Kuethe, and M. Morehart. 2015. "The impact of decoupled payments on US cropland values." *Agricultural Economics* 46:643-652.

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.
Karlen, D.L., et al. 2006. "Crop rotation effects on soil quality at three northern corn/soybean belt locations." *Agronomy journal* 98:484-495.
Kiel, K.A., and M. Williams. 2007. "The impact of Superfund sites on local property values: Are all sites the same?" *Journal of urban Economics* 61:170-192.
Kraenzel, D.G., et al. 1998. "Industrial hemp as an alternative crop in North Dakota." *Agricultural Economics Report* 402.
Leteinturier, B., et al. 2006. "Adaptation of a crop sequence indicator based on a land parcel management system." *Agriculture, Ecosystems & Environment* 112:324-334.
Lichtenberg, E. 1989. "Land quality, irrigation development, and cropping patterns in the northern high plains." *American Journal of Agricultural Economics* 71:187-194.
Livanis, G., et al. 2006. "Urban sprawl and farmland prices." *American Journal of Agricultural Economics* 88:915-929.
Long, J.A., et al. 2014. "Changes in field-level cropping sequences: Indicators of shifting agricultural practices." *Agriculture, Ecosystems & Environment* 189:11-20.
Louviere, J.J., D.A. Hensher, and J.D. Swait. 2000. *Stated choice methods: analysis and applications*: Cambridge university press.
Martinez, A., and D.E. Maier. 2014. "Quantification of Biomass Feedstock Availability to a Biorefinery Based on Multi-Crop Rotation Cropping Systems Using a GIS-Based Method." *Biological Engineering Transactions* 7:3-16.
Matis, J., et al. 1985. "A Markov chain approach to crop yield forecasting." *Agricultural Systems* 18:171-187.
McFadden, D. 1973. "Conditional logit analysis of qualitative choice behavior."
McFadden, D. 1980. "Econometric models for probabilistic choice among products." *Journal of Business*:S13-S29.
McGranahan, D.A. "Natural amenities drive rural population change."
Messer, K.D., et al. 2006. "Can stigma explain large property value losses? The psychology and economics of Superfund." *Environmental and Resource Economics* 33:299-324.
Muller, M.R., and J. Middleton. 1994. "A Markov model of land-use change dynamics in the Niagara Region, Ontario, Canada." *Landscape ecology* 9:151-157.
Newhouse, J.P. 1987. "Health economics and econometrics." *The American economic review* 77:269-274.
Newman, C., M. Henchion, and A. Matthews. 2003. "A double-hurdle model of Irish household expenditure on prepared meals." *Applied Economics* 35:1053-1061.
Nickerson, C., et al. 2012. "Trends in US farmland values and ownership."
Nickerson, C.J., and L. Lynch. 2001. "The effect of farmland preservation programs on farmland prices." *American Journal of Agricultural Economics* 83:341-351.
Oborne, M. 2010. "The bioeconomy to 2030: designing a policy agenda." *Organisation for Economic Cooperation and Development. The OECD Observer*:35.

Osman, J., J. Inglada, and J.-F. Dejoux. 2015. "Assessment of a Markov logic model of crop rotations for early crop mapping." *Computers and Electronics in Agriculture* 113:234-243.

Palmquist, R.B. 2005. "Property value models." *Handbook of environmental economics* 2:763-819.

Palmquist, R.B., and L.E. Danielson. 1989. "A hedonic study of the effects of erosion control and drainage on farmland values." *American Journal of Agricultural Economics* 71:55-62.

Park, J., and G.C. Davis. 2001. "The theory and econometrics of health information in cross-sectional nutrient demand analysis." *American Journal of Agricultural Economics* 83:840-851.

Paton, L., et al. (2014) "Multinomial logistic regression on Markov chains for crop rotation modelling." In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems.* Springer, pp. 476-485.

Plantinga, A.J., T. Mauldin, and D.J. Miller. 1999. "An econometric analysis of the costs of sequestering carbon in forests." *American Journal of Agricultural Economics* 81:812-824.

Plantinga, A.J., R.N. Lubowski, and R.N. Stavins. 2002. "The effects of potential land development on agricultural land prices." *Journal of urban Economics* 52:561-581.

Plourde, J.D., B.C. Pijanowski, and B.K. Pekin. 2013. "Evidence for increased monoculture cropping in the Central United States." *Agriculture, Ecosystems & Environment* 165:50-59.

Porter, P.M., et al. 1997. "Environment affects the corn and soybean rotation effect." *Agronomy journal* 89:442-448.

Ritter, M., et al. 2019 "Revisiting The Relationship Between Land Price And Parcel Size." FORL and Working Paper. https://doi.org/10.18452/19877.

Robison, L.J., D.A. Lins, and R. VenKataraman. 1985. "Cash rents and land values in US agriculture." *American Journal of Agricultural Economics* 67:794-805.

Rosen, S. 1974. "Hedonic prices and implicit markets: product differentiation in pure competition." *The journal of political economy*:34-55.

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

Savage, T. 2011. "Forecasting land use from estimated markov transitions."

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

Stern, A.J., P.C. Doraiswamy, and E.R. Hunt. 2012. "Changes of crop rotation in Iowa determined from the United States Department of Agriculture, National Agricultural Statistics Service cropland data layer product." *Journal of Applied Remote Sensing* 6:063590-063590.

Taylor, H.M., and S. Karlin. 2014. *An introduction to stochastic modeling*: Academic press.

Thornton, P.K., and P.G. Jones. 1998. "A conceptual approach to dynamic agricultural land-use modelling." *Agricultural Systems* 57:505-521.

Troffaes, M., and L. Paton (2013) "Logistic regression on Markov chains for crop rotation modelling." In., Society for Imprecise Probability: Theories and Applications (SIPTA).

U.S. Department of Agriculture, National Agricultural Statistics Service. 2016. Land Values 2015 Summary: August 2016. http://usda.mannlib.cornell.edu/usda/nass/AgriLandVa//2010s/2016/AgriLandVa-08-05-2016.pdf (accessed 20 August 2017).

Vasquez, O., J.R. Nelson, and J.R. Hamilton. 2002. "Regression analysis to determine the effects of land characteristics on farmland values in South-Central Idaho." *Journal of the American Society of Farm Managers and Rural Appraisers* 65:69-77.

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

Vogel, F.A., and G.A. Bange. 1999. "Understanding USDA crop forescasts."

Wagner, J., and S. Hanna. 1983. "The effectiveness of family life cycle variables in consumer expenditure research." *Journal of Consumer Research* 10:281-291.

Wasson, J.R., et al. 2013. "The effects of environmental amenities on agricultural land values." *Land Economics* 89:466-478.

Wu, J., et al. 2004. "From microlevel decisions to landscape changes: an assessment of agricultural conservation policies." *American Journal of Agricultural Economics* 86:26-41.

Yost, M.A., et al. 2014. "Alfalfa stand length and subsequent crop patterns in the upper midwestern United States." *Agronomy journal* 106:1697-1708.

Zhang, F., et al. 2008. "Modeling fresh organic produce consumption with scanner data: a generalized double hurdle model approach." *Agribusiness* 24:510-522.

Zhang, W., and C.J. Nickerson. 2015. "Housing market bust and farmland values: Identifying the changing influence of proximity to urban centers." *Land Economics* 91:605-626.

VITA

GWANSEON KIM

EDUCATION

- M.S., Agricultural Economics, Mississippi State University, MS, May 2011
- B.A., Business Information System, Mississippi State University, MS, May 2009

RESERCH & PROFESSIONAL POSITIONS

- Research Assistant, Department of Agricultural Economics, University of Kentucky, July 2014-2019.
- Research Assistant, Department of Agricultural Economics, Mississippi State University, Aug 2009-2011.
- Teaching Assistant, Department of Agricultural Economics, University of Kentucky, 2015-2016.
- Lecturer, Department of Agricultural Economics, University of Kentucky, 2016-2019.

REFEREED JOURNAL ARTICLES

- **Kim, G.,** J.H. Seok, T.B. Mark, and M.R. Reed. "The Price Relationship between Organic and Non-Organic Vegetables in the U.S.: Evidence from Nielsen Scanner Data." *Applied Economics* (2018): 1-15.
- **Kim, G.,** J.H. Seok, and T.B. Mark. "New Market Opportunities and Consumer Heterogeneity in the U.S. Organic Food Market." *Sustainability* 10.9 (2018): 3166. doi:10.3390/su10093166.
- Seok, J.H., H. Moon, **G. Kim**, M.R. Reed. "Is Aging the Important Factor for Sustainable Agricultural Development in Korea? Evidence from the Relationship between Aging and Farm Technical Efficiency." *Sustainability* 10.7 (2018): 1-15.
- Seok, J.H., **G. Kim**, M.R. Reed, and S. Kim. "The Impact of Avian Influenza on the Korean Egg Market: Who Benefited?" *Journal of Policy Modeling* 40.1 (2018): 151-165.
- Pak, T.Y. and **G. Kim**. "The impact of Medicare Part D on Cognitive Functioning at Older Ages." *Social Science & Medicine* 193 (2017): 118-126.
- **Kim, G.** and T.B Mark. "Impacts of corn price and imported beef price on domestic beef price in South Korea." *Agricultural and Food Economics* 5.1 (2017): 5.
- Seok, J.H., **G. Kim**, and T.B Mark. "The Impact of Minimum Wage on Food Away from Home Expenditure Using Structural Equation Model." *International Journal of Food and Agricultural Economics* 5.2 (2017): 45.
- **Kim, G.** and Y. Zheng. "U.S. Nonalcoholic Beverage Demand: Evidence from AIDS Model with Dynamic Effect." *Journal of Agribusiness* 35 (2017): 1-14.

- **Kim, G.,** D.R. Petrolia, and M.G. Interis. "A Method for Improving Welfare Estimates from Multiple-Referendum Surveys." *Journal of Agricultural and Resource Economics* 37.2 (2012): 289-300.

OTHER PUBLICATIONS

- **Kim, GwanSeon**. “The What and Why of Organic Produce Prices.” *American Vegetable Grower*, October 2018: 4-6.
- Petrolia, D.R., M.G. Interis, M.K. Hidrue, J. Hwang, **G. Kim**, R.G. Moore. “America’s Wetland? A National Survey of Willingness to Pay for Restoration of Louisiana’s Coastal Wetlands.” Final Project Report, Department of Agricultural Economics, Mississippi State University, April 2012.

WORKING PAPERS

- **Kim, G.,** M. Nemati, S.C. Buck, and T.B. Mark. “Recovering forecast distributions of Crop Composition: Method and Application to Kentucky Agriculture.” *Journal of Agricultural & Applied Economics* [Revise & Resubmit]
- **Kim, G.,** J. Schieffer, and T.B. Mark. "Do Superfund Sites Affect Local Property Values? Evidence from a Spatial Hedonic Approach." [Under Review]
- **Kim, G.,** S.K. Choi, and J.H. Seok. “Does Biomass Energy Consumption Reduce Total Energy CO_2 Emissions in the U.S.?” [Under Review]
- Seok, J.H. and **G. Kim**. “Asymmetric Price Transmission from Crude Oil to Domestic Gasoline in South Korea.” [Under Review]
- Pak, T.Y. and **G. Kim**. “Food Stamps, Food Insecurity, and Health Outcomes among Elderly Americans.” [Under Review]
- **Kim, G.** and T.B. Mark. “Markets for Industrial Hemp are Smokin’, Despite Legal-Limbo Status” [In Progress]
- **Kim, G.** and T.B. Mark. “Factors Affecting Heterogeneous Agricultural Land: The Case of Kentucky.” [In Progress]
- Pates, N., **G. Kim**, and T.B. Mark. “Mid-Season Production Throttles: An Analysis of the Intra-Seasonal Field Crop Yield-Price Relationship in the US.” [In Progress]
- Seok, J.H., **G. Kim**, and T.B. Mark. “Does the Agricultural Growth Lead the Methane Emissions? Evidence from 34 Major Meat Production Countries.” [In Progress]