Article

Recovering Forecast Distributions of Crop Composition: Method and Application to Kentucky Agriculture

GwanSeon Kim 1,*, Mehdi Nemati 2, Steven Buck 3, Nicholas Pates 3 and Tyler Mark 3

1 College of Agriculture, Arkansas State University, Jonesboro, AR 72467, USA
2 School of Public Policy, University of California, Riverside, CA 92521, USA; mehdi.nemati@ucr.edu
3 Department of Agricultural Economics, University of Kentucky, Lexington, KY 40546, USA;
    steven.buck@uky.edu (S.B.); nicholas.pates@uky.edu (N.P.); tyler.mark@uky.edu (T.M.)

* Correspondence: gkim@astate.edu; Tel.: +1-870-972-2263

Received: 9 January 2020; Accepted: 3 April 2020; Published: 6 April 2020

Abstract: This paper proposes a novel application of the multinomial logit (MNL) model using Cropland Data Layer and field-level boundaries to estimate crop transition probabilities, which are used to generate forecast distributions of total acreage for five major crops produced in the state of Kentucky. These forecasts distributions have a wide range of applications that, besides providing interim acreage estimates ahead of the June Acreage Survey, can inform the ability of producers to incorporate new crops in the land-use rotation, investments in location-specific capital and input distribution as well informing the likelihood of adverse water quality events from nutrient run-off.

Keywords: forecast distribution; multinomial logit model; simulation; transition probability

1. Introduction

Improvements in crop production forecasts and forecasting methods have been a focus of agricultural economics research for decades because production forecasts directly influence year-to-year local, state, regional, national, and international economies. Crop production assessments also provide important implications for agribusiness and food management [1]. On a macro level, understanding the determinants of crop acreage helps with the identification and management of the demand and supply of crop production [2]. For example, merchandizers rely on crop supply and demand estimates prepared by both public and private organizations [3]. Moreover, estimates and forecasts of acreage can have a significant impact on futures prices and market volatility [4], as well as market participants [5]; thus, impacting farm income and investment in agriculture. As a consequence, the release of forecasts is highly anticipated market events [2,4].

The National Agricultural Statistics Service (NASS) of the U.S. Department of Agriculture (USDA) is the primary provider of public information on crop acreage forecasts [4]. According to Adjemian and Smith [6], 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, weather forecasts, and expected market development from the NASS and Interagency Commodity Estimates Committees (ICEC). In addition to the NASS estimates, private expectations data by Conrad Leslie and Informa Economics (previously Sparks Companies) have been used for crop acreage forecasts, especially for corn, soybeans, and wheat [5]. These 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.
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 Acreage Survey [7]. The survey used to estimate acreage is based on a random sample of farm operators instead of all operators [4]; thus, the estimated acreage is subject to sampling error.

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, they do not reflect uncertainty in these forecasts, which is a consideration of agricultural decision-makers. As an example, the 2019 June Acreage Report released by USDA NASS has been highly criticized by analysts and farm organizations for overstating the number of corn acres planted [8]. Furthermore, USDA has now indicated that as a result of the unreliable estimates in this report, they resurveyed the entire US Corn Belt in August 2019. By shifting to the forecasting of acreage distributions instead of mean acreage, uncertainty in the forecast values such as minimum and maximum forecasted acreages are incorporated.

The main objectives of this paper are three-fold. First, for the purpose of forecasting, we estimate the conditional transition probabilities for crop choice with a multinomial logit (MNL) model using the Cropland Data Layer (CDL) and field level boundaries. While others have taken advantage of these new data sources to estimate a discrete choice model of crop choice, Hendricks, et al. [9] as a prominent example, such analyses have been built around empirical objectives of description, identification of causal effects or testing a theory of agricultural land-use, which are distinct from our empirical objective of forecasting land-use distributions.

For our second objective, we use conditional transition probabilities estimated from the MNL model to simulate the distribution of forecasted total acreages for the five largest crops produced in the state of Kentucky. These forecast distributions can be used in a variety of ways. The June Acreage Survey, while highly accurate for areas surveyed, is not ground-truthed. The quality of NASS crop acreage and production estimates depends on a high level of participation by agricultural producers [10]. 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 [10,11].

We model using data from the CDL, which is based on comprehensive satellite imagery for the areas under study and not subject to the sampling error present in the June Acreage Survey, although the satellite images are still subject to measurement error [12]. Our mean forecasts directly compare to the June Acreage Survey and could provide supplementary acreage information earlier in the season. Without shoe-leather, we cannot ground-truth either our estimates or those of the June Acreage Survey. It is also the case that our forecasts for land-use in the following year are in advance of the June Acreage Survey. The essential data required for our approach is the CDL, which becomes available early in the year; for example, the 2018 CDL became available on 15 February 2019. In the future, satellite imagery may be feasibly measure planted acreage earlier than the current CDL, which would make CDL-type estimates available even earlier than mid-February. Therefore, our approach can be harnessed to generate interim acreage estimates prior to the release of the June Acreage Survey that is released by the end of June. The June Acreage is the first report released by USDA after the March Planting Intentions and is a barometer for crop expectations.

Third, our approach generates a distribution of potential land-use outcomes that can serve as inputs into production management and policy decisions. For example, land-use forecast distributions also speak to the adaptability of farmers within Kentucky to switch to an alternative crop, such as hemp, that has over 57,000 acres of production licensed for 2019. Wider distributions suggest more flexibility in crop composition mix. Agricultural land-use forecast distributions can be used to quantify uncertainty affecting investments in location-specific capital (e.g., where to build new distilleries in Kentucky) and where to store inventory for supply inputs such as fertilizer. Forecast distributions may also be used to provide insight into the likelihood of soil erosion, nitrogen run-off and other events adversely affecting soil and water quality. Specifically, each land-use can be linked to an expected soil or water quality outcome (e.g., nutrient run-off); thus, we can use these links to generate distributions of soil
or water quality outcomes. It is also the case, that our approach could be employed to generate distributions of land-use forecasts several years into the future, which is another advantage relative to the existing June Acreage Survey estimates.

In summary, our work represents an exercise that (i) makes new use of the Cropland Data Layer and field-level boundaries for generating land-use forecasts, (ii) proposes a novel application of the multinomial logit model to generate forecast distributions of total acreage for the five largest crops produced in the state of Kentucky and (iii) provides a discussion on how to harness the agricultural land-use forecast distributions to inform policy-discussions related to investments in location-specific capital, distribution of supply inputs, soil and water quality management as well as viability of new crops such as hemp becoming a significant part of a state’s crop portfolio.

The remainder of the paper is structured as follows: Section 2 discusses related literature, Section 3 describes and presents data. Section 4 explains the empirical models, including the multinomial logit model and a first-order Markov chain approach. Section 5 discusses the analysis and presents the results. We conclude the paper in Section 6 and offer policy implications and areas for future research.

2. Literature Review

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 [13]. Farmers commonly practice crop rotation because its advantages offset its disadvantages. Crop rotation benefits include increased yield [14,15], improved soil fertility [16–18], reduced greenhouse gas emissions [19], and reduced economic risk by having more than one crop as a potential income source [13]. 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 the local agricultural system [20,21].

A MNL method, developed by McFadden [22], is employed in this study to model the crop choice behavior. Using this method, we can model a farmer’s choice of a crop among different and exclusive alternatives (i.e., corn, soybean, tobacco, wheat, and alfalfa) and estimate Markov transition probabilities for these crops. Later in Section 4.1, we explain the MNL method in detail. The MNL model has been widely used in agricultural-related studies especially for modeling land use [23–26]. Hardie and Parks [24] employ the MNL model into the land use decision by incorporating heterogeneous land quality in the southeastern U.S. Plantinga, Mauldin and Miller [25] simulate carbon sequestration based on estimates of land use share in Maine, South Carolina, and Wisconsin by utilizing the MNL model. Wu, Adams, Kling and Tanaka [26] 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, Flores-Lagunes and Guci [23] use the MNL model by incorporating spatial dependence in Medina County, Ohio, for the determinants of land-use choices. Paton, et al. [27] 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. [28] 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 the forecasting mean yields. Following the literature in crop yield distribution forecasting, we use the first-order Markov Chain approach to predict crop acreage distributions by accounting for the dynamics in the analysis with specifying the year-to-year transitions between five crops. To the best of our knowledge, our approach has not been employed previously to forecast crop acreage distributions, though there are several relevant applications to inform producers and policymakers.
3. Data

The primary source of information for our crop choice data is the Cropland Data Layer (CDL) produced by NASS. The CDL was initiated in early 1997 on a limited basis to provide annual geospatial content to customers who were interested in annual cropland cover updates. Starting in 2008, the CDL provides coverage for the lower 48 states in the U.S. The CDL provides a comprehensive, raster-formatted, and geo-referenced imagery for crop-specific land cover classification to identify field crop types accurately and geospatially [29]. Note that 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. The CDL includes crop or land use classification codes, which are assigned to each pixel and classified by NASS using data from satellite sensors [9,16]. The use of the CDL data to date has been limited but has received more attention recently to study farmer’s behavior regarding crop choice [9,16,18,21,30,31]. We use the CDL observations for Kentucky from 2010–2015. We focus our analysis on five main crops: corn, soybeans, tobacco, wheat, and alfalfa. These are the top five-row crops in Kentucky based on acreage (see Figure A1 in the Appendix B). This will provide an idea of how much areas these crops occupy in Kentucky Next, we employ the Common Land Unit (CLU) boundaries, obtained from the GeoCommunity to identify field boundaries. 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 includes a permanent, contiguous boundary, common land cover, and land management.

To construct a field-level crop choice dataset we used the following steps: First, to remove non-agricultural fields we overlay the CLU with the 2011 National Land Cover Dataset (NLCD) 2011, which is the most recent national land cover product, produced by the Multi-Resolution Land Characteristics (MRLC). 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 a raster based on the majority of adjacent cells; this removes mis-specified (i.e., spurious) cells and to smooth the raster. 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. Table 1 shows the total number of 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. In Appendix A, we describe the data merging and cleaning process in detail. In 2015, soybeans, corn, alfalfa, tobacco, and wheat acreage represented 42%, 36%, 1%, 0.6%, and 0.4% of total acres in Kentucky, respectively. As illustrated in Figure A1 in Appendix B, the majority of corn and soybeans are planted in western Kentucky.

A data limitation is that 28 (out of 120) counties in Kentucky are excluded from this study because the CLU data was not available. Table 2 shows the missing acres in percentage compared to the original CDL data. Based on Table 2, we are losing more data for tobacco and alfalfa compared to corn, soybeans, and wheat. Figure A2 in Appendix B shows the locations of excluded counties.

Crop choice decisions by farmers are partially dependent on the weather (e.g., precipitation and temperature) observed in a growing season and, therefore, are predictors of land-use [32]. We obtained precipitation and temperature data from the Parameter-elevation Regressions on Independent Slopes Model (PRISM). For the weather variables, we use average precipitation in April, May, and June, the average temperature in June, July, and August months of each year. Soil quality is another predictor of crop choice behavior by farmers; thus, 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). Finally, we obtain cross-sectional National Elevation data (30-meter resolution) from the Geospatial Data Gateway provided by USDA-NRCS to calculate the elevation and slope. All of the data, which are precipitation, temperature, slope, elevation, and soil textures, are spatially joined based on the unique field ID. Table 3 shows the summary statistics of the data used in this study. As shown in Table 3, the mean acreage of these fields is 5.97 acres, and their soil composition is 66.3% silt, 20.6% clay,
and 11.8% sand, on average. These fields represent 92 (out of 120) counties in Kentucky. Our sample covers 92 counties accounting for 32,149.43 square miles which corresponds to almost 80% of the total land in Kentucky. In addition, the average monthly temperature and precipitation are 29.94 °C and 143.68 mm, respectively.

Table 1. Total number of fields in the sample by crop class by year.

| Crop     | 2010  | 2011  | 2012  | 2013  | 2014  | 2015  | Sum      |
|----------|-------|-------|-------|-------|-------|-------|----------|
| Corn     | 105,266 | 100,060 | 123,638 | 112,930 | 113,079 | 113,662 | 668,635  |
|          | (33.70) | (32.03) | (39.58) | (36.15) | (36.20) | (36.39) |          |
| Soybean  | 98,311  | 79,480  | 76,268  | 79,492  | 99,343  | 129,684 | 562,578  |
|          | (31.47) | (25.44) | (24.41) | (25.45) | (31.80) | (41.51) |          |
| Tobacco  | 456    | 298    | 734    | 1056   | 1317   | 1730   | 5591     |
|          | (0.15) | (0.10) | (0.23) | (0.34) | (0.42) | (0.55) |          |
| Wheat    | 276    | 254    | 267    | 344    | 203    | 1232   | 2576     |
|          | (0.09) | (0.08) | (0.09) | (0.11) | (0.06) | (0.39) |          |
| Alfalfa  | 1742   | 2936   | 1593   | 1537   | 2168   | 3136   | 13,112   |
|          | (0.56) | (0.51) | (0.49) | (0.69) | (1.00) |        |          |
| Other    | 106,313 | 129,336 | 109,864 | 117,005 | 96,254 | 62,920 | 621,692  |
|          | (34.03) | (41.40) | (35.16) | (37.45) | (30.81) | (20.14) |          |
| 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 a year. These numbers are calculated as dividing the number of fields for each crop by the total number of fields in each year (last row in Table 1).

Table 2. Percentage of missing acres in our analysis based on missing counties in the Common Land Unit (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 3. Summary of data used for the estimation of conditional probabilities.

| Variable          | Mean   | SD   | 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 parenthesis. Acres indicate a number of acreages planted, and zero acreage refers to another choice. SD represents the standard deviation.
4. Empirical Model

4.1. Multinomial Logit Model

This study employs the MNL model developed by McFadden [22] to estimate Markov transition probabilities for the five primary crops in Kentucky: corn, soybean, tobacco, wheat, and alfalfa. In our setting, the farmer chooses one alternative among different and exclusive alternatives. The random utility model (RUM) provides the theoretical economic justification for using an MNL when modeling choices. Under the RUM framework the chosen the alternative is determined by the utility level, \( U_{ij} \), each choice \( j \), provides the farmer \( i \). It follows that

\[ U_{ij} = x'_{ij} \beta_j + \epsilon_{ij} = V_{ij} + \epsilon_{ij} \quad (1) \]

where \( i = 1, \ldots, N; \ j = 1, \ldots, m; \ V_{ij} \) is a deterministic component that depends on regressors and unknown parameters, \( \epsilon_{ij} \) is the unobserved component. The alternative providing the highest level of utility will be chosen; in other words, alternative \( j \) is chosen if and only if \( \forall k \neq j \ U_{ij} > U_{ik} \).

Suppose a farmer chooses at least one crop among competing alternatives. The farmer will choose corn if the utility by choosing corn is higher than the utility by choosing soybeans (i.e., \( U_{\text{Corn}} > U_{\text{Soybean}} \)). 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 \( y_t = j \) if alternative \( j \) provides the highest utility of the alternative, and the general expression of the probability of choosing alternative \( j \) can be defined as:

\[ P_{ij} = P(y_t = j) = P(U_{ij} > U_{ik}\mid x, \forall k \neq j) = P(\epsilon_{ik} - \epsilon_{ij} \leq x'_{ij} \beta_j - x'_{ik} \beta_k \mid x, \forall k \neq j) \quad (2) \]

According to Croissant [33], the MNL model generally assumes that error terms are independently and identically distributed (IID). With this assumption, Equation (2) can be shown that the choice probabilities are

\[ P_{ij} = P(y_t = j) = \frac{\exp(x'_{ij} \beta_j)}{\sum_{k=1}^{m} \exp(x'_{ik} \beta_k)} \quad (3) \]

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 predict crop choice behavior of Kentucky farmers. This study also considers and includes one more alternative, called “other.” The choice of other represents that farmers do not plant any of five major crops in this field during our study periods or other crops. The other category includes fallow, oats, barley, grains sorghum, and double-crop beans. By setting \( \beta_j = 0 \) for the choice of other, the set of coefficients corresponding to each outcome are estimated as Equation (4). Note that 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, the greater number does not imply better outcome compared to the smaller number. In addition, the outcome of no production (i.e., the choice of other) is our base outcome: \( 1=0 \).

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

4.2. Markov Chain Approach

The Markov chain approach has been widely used in land use studies [34–39] as well as agricultural-related studies especially for studying the crop rotation behavior [27,28,40–43]. Based on Taylor and Karlin [44] and Savage [38], 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 predictive of the probability of any specific future behavior if the current state is known. The Markov property, in general, can be defined as the following:

$$\Pr[X_{t+1} = j | X_0 = i_0, \ldots, X_{t-1} = i_{t-1}, X_t = i] = \Pr[X_{t+1} = j | X_t = i]$$  \hspace{1cm} (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 $P^{t+1}_{ij}$. The Markov chain is stationary if the one-step transition probabilities are not a function of the time variable $t$. Considering a stationary Markov chain, the one-step transition probabilities are re-written as

$$P^{t+1}_{ij} = P_{ij} = \Pr[X_{t+1} = j | X_t = i]$$  \hspace{1cm} (6)$$

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

$$P[X_t = j | X_{t-1} = i_{t-1}, X_{t-2} = i_{t-2}, \ldots] = P[X_t = j | X_{t-1} = i_{t-1}]$$  \hspace{1cm} (7)$$

With a 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, and the transition probability matrix has $6 \times 6$ with elements, which determine the probability of transitioning from a row outcome to a column outcome: the row outcome represents a cropping choice among five crops in the previous year, and the column outcome represents the crop choice in the current year.

5. Empirical Results

5.1. MNL Results and Transition Probabilities

This study implements several hypotheses tests for measure fit. First, 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 the results of the LR and Wald tests, we reject the null hypothesis that all coefficients associated with given variables are zero, implying any variables cannot be dropped from the model since independent variables have a significant impact across all crop choices. Second, 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 no evidence to support combining outcome variable categories.

Table 4 presents the estimated results of the MNL model; in a third test, we see the chi-square and p-values indicate that the model as a whole is statistically significant. The results are consistent with the expected crop rotation results for all five crops used in this study. For example, in the case of corn, it is expected that there is a higher likelihood of 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 5. Based on the marginal effects in Table 5, it shows that if corn was planted in the previous year, then compared to other choice category, it is associated with corn being 6.1% less likely, soybean being 27.7% more likely, tobacco being 0.002% more likely, wheat is 0.1% more likely, and alfalfa is 0.01% less likely to be planted. In Table 5, we do not report the marginal effect of the choice of ‘other’. In general, the marginal effects sum up to zero.
### Table 4. Conditional multinomial logit model results.

| Choice lag | Corn      | Soybean    | Tobacco    | Wheat      | Alfalfa    |
|------------|-----------|------------|------------|------------|------------|
| Corn       | 0.085 *** | 2.415 ***  | 0.463 ***  | 2.327 ***  | -0.847 *** |
|            | (0.0055)  | (0.0081)   | (0.0587)   | (0.0894)   | (0.0608)   |
| Soybean    | 2.172 *** | 2.935 ***  | 1.369 ***  | 2.198 ***  | 0.171 ***  |
|            | (0.0062)  | (0.0091)   | (0.0630)   | (0.1025)   | (0.5599)   |
| Tobacco    | 0.334 *** | 1.328 ***  | 4.715 ***  | 2.060 ***  | -0.575 *** |
|            | (0.0406)  | (0.0500)   | (0.0670)   | (0.3378)   | (0.4057)   |
| Wheat      | 1.522 *** | 1.501 ***  | 1.620 ***  | 3.456 ***  | 2.143 ***  |
|            | (0.0494)  | (0.0769)   | (0.3991)   | (0.3018)   | (0.1985)   |
| Alfalfa    | -1.061 ***| -0.213 *** | -1.205 *** | 0.254      | 3.947 ***  |
|            | (0.0399)  | (0.0571)   | (0.4614)   | (0.4251)   | (0.0376)   |

**Observations:** 1,561,820
**McFadden R-Square:** 0.249
**Log Likelihood:** -1,285,868

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

### Table 5. Marginal effects from conditional multinomial logit model results.

| Choice lag | Corn      | Soybean    | Tobacco    | Wheat      | Alfalfa    |
|------------|-----------|------------|------------|------------|------------|
| Corn       | -0.061 ***| 0.277 ***  | 0.00002 *  | 0.001 ***  | -0.0001 ***|
|            | (0.0009)  | (0.0008)   | (0.00001)  | (0.00006)  | (0.0000)   |
| Soybean    | 0.347 *** | 0.169 ***  | 0.00001    | 0.0002 *** | -0.0001 ***|
|            | (0.0011)  | (0.0008)   | (0.00001)  | (0.00003)  | (0.0000)   |
| Tobacco    | 0.033 *** | 0.084 ***  | 0.016 ***  | 0.001 **   | -0.0001 ***|
|            | (0.0077)  | (0.0051)   | (0.0016)   | (0.0003)   | (0.0002)   |
| Wheat      | 0.031 *** | 0.049 ***  | 0.0003     | 0.002 ***  | 0.0004 *** |
|            | (0.0110)  | (0.0057)   | (0.0002)   | (0.0008)   | (0.0001)   |
| Alfalfa    | -0.146 ***| -0.001     | -0.0001 ***| 0.0001     | 0.008 ***  |
|            | (0.0037)  | (0.0022)   | (0.00003)  | (0.0001)   | (0.0007)   |

Notes: Robust standard errors are reported in parentheses. 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.

As another measure of model performance, McFadden’s pseudo $R$-squared commonly used for summarizing in-sample model fit. Based on Louviere, et al. [45], in-sample model fit is considered to be extremely good if the value of the McFadden’s $R$-squared is between 0.2 and 0.4. Domenech and Mcfadden [46] argue this range is equivalent to 0.7 to 0.9 for a linear regression model; our model has a pseudo $R$-squared of 0.25 indicating a high degree of in-sample explanatory power.

We also conduct an out-of-sample validation test for the predicted probabilities from the MNL model. In other words, how accurately does the estimated model predict the crop choice? For the validation test, we employ the following steps. First, we select 10% of the data using a random sampling process, which we call our test sample. Second, we estimate the MNL model with the remaining dataset (90% of data, also called training set), and then predict the crop choice probabilities for each observation in the test sample. Third, we assign crops based on the highest predicted probabilities for each field. For example, if a field has predicated probabilities of 0.55, 0.20, 0.10, 0.5, 0.5, and 0.5 for corn, soybeans, tobacco, alfalfa, wheat, and other, respectively then corn (having the highest probability
of being planted) is assigned to the field. Finally, we compare these predictions with actual crop choice in each field and calculate the acreage accuracy. Based on the out-of-sample validation exercise, 60.7% were correctly predicted. We also resize the test sample sizes by 20% and 30%, and we find that probabilities are correctly predicted by 61.4% and 60.7%, respectively.

As a summary of the MNL model results presented in Table 4, we calculate average conditional predicted probabilities or Markov transition probabilities over our study period. Table 6 shows average transition probabilities based on the annual predicted transition probabilities from 2010 through 2015. Based on Table 6, this study finds that if corn is planted in year \(t-1\), then the average predicted transition probability between 2010 and 2015 is 0.411 probability that corn will be planted in year \(t\). As a summary, average crop rotation probabilities between corn to soybean and soybean to corn from the year \(t-1\) to the year \(t\) are 0.239 and 0.309, 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 soil fertility but also protecting the environment from the nitrogen runo

Table 6. Conditional transition probabilities.

| From:       | Other | Corn | Soybean | Tobacco | Wheat | Alfalfa |
|-------------|-------|------|---------|---------|-------|---------|
| Other       | 0.542 | 0.255| 0.193   | 0.002   | 0.001 | 0.007   |
| Corn        | 0.343 | 0.411| 0.239   | 0.002   | 0.001 | 0.003   |
| Soybean     | 0.348 | 0.309| 0.336   | 0.003   | 0.002 | 0.003   |
| Tobacco     | 0.473 | 0.282| 0.205   | 0.031   | 0.002 | 0.007   |
| Wheat       | 0.403 | 0.267| 0.317   | 0.005   | 0.002 | 0.006   |
| Alfalfa     | 0.640 | 0.141| 0.072   | 0.003   | 0.001 | 0.143   |

Notes: Estimated conditional transition probabilities are based on the conditional multinomial logit model. For example, transition probability from corn to corn is 0.411. In this table, “Other” include fallow, oats, barley, grains, sorghum, double-crop beans, etc.

5.2. Simulation Exercise

The objective of this section is to generate a forecast distribution of total acreage for crop \(I\) in the year \(t\) using information up to year \(t-1\) (e.g., predicting crop composition in 2016 based on using data up to 2015 to estimate the MNL model). For this purpose, we follow three steps. First, we use an MNL specified in the previous section to estimate the probability of a farmer choosing crop \(I\) for field \(j\). Using this method, we generate a probability distribution for each field \(j\). Second, we make a single draw from each field’s distribution (e.g., field 1 turns up corn, field 2 turns up corn, field 3 turns up soybeans and so on for each field). Based on this single draw for all fields and knowing the acreage corresponding to each field, we are able to sum up the aggregate (forecasted) acreage corresponding to each crop—this is one realization of the aggregate composition of acreage (for one draw). Third, we complete 1000 such draws; thus, generating 1000 forecasts of aggregate acres by crop. It is exactly these 1000 draws that we use to generate our forecast distributions by crop type.

We apply this method to Kentucky fields and predict the expected crop composition in 2016 based on our MNL model results. Figure 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. 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, which makes sense given growing and contractual considerations. 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 fairly stable over time. Tobacco is primarily produced via a production contract. This creates a situation in which the acreage will be quite stable from year to year. Likewise, for wheat, Kentucky is home to a significant milling and distilling industries that have relatively stable contracts with producers. Lastly, this leaves corn and soybean acres to be the primary acreages in the state that shift to fill in the gaps. These forecast distributions demonstrate this phenomenon.

![Density plots for each crop in 2016](image)

**Figure 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. SD represents the standard deviation, and CV represents a coefficient of variation.)
As a summary of results, 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 soybeans, 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.

5.3. Discussion of Production and Policy Implications

The forecast distributions we generate can inform both short and long-run investment decisions. In the short run, fertilizer suppliers must decide how much fertilizer to have in inventory and to where they should anticipate sending fertilizer supply [47]. Location-specific demand for fertilizer is uncertain and our forecast distribution reflects this uncertainty. For example, in response to uncertainty, fertilizer suppliers may have ready buffer fertilizer supplies where forecast distributions are wider.

In the longer term, investments in location-specific capital will benefit from the knowledge of forecast distributions. For a historical example from the recent past, an ethanol plant will generally have a higher Return on Investment (ROI) when there is a high degree of certainty for expected crop production, in the case of an ethanol plant - expected corn production. While there is less discussion of new ethanol plants in Kentucky, the distillery industry continues to grow in the state and requires investment in new distilleries, another example of location-specific capital [48]. Knowledge of the distribution of corn and wheat compared to soybean and new crops such as hemp, which benefits from flex in land use composition, can support investment decisions in the industry. Note that Kentucky is ranked the third largest state in the U.S. to grow hemp. Total areas of hemp grown in 2017 and 2018 were 3271 acres and 6700 acres, respectively. In addition, growing industrial hemp as an agricultural commodity in the U.S. was legalized based on the 2018 Farm Bill, suggesting more room for growth.

Our forecast distributions may also offer insight into nutrient load management and water quality outcomes. For example, with parcel-specific information on soil characteristics, land slope and proximity to waterways, then predictions of nutrient run-off can be generated for crop choice under normal and adverse weather conditions. Such information may be combined with our forecast distributions to generate nutrient run-off distribution for both normal and adverse weather conditions. Further, these distributions may be combined with water quality standards to calculate exceedance probabilities on the likelihood that nutrient loads in local waterways exceed water quality standards under normal and adverse weather conditions. To enrich this analysis for specific applications, qualitative information could be collected for all the farmers in the region on production practices, soil test values, and other demographic information to further enhance the prediction ability of the model.

6. Concluding Remarks

This study proposes a novel application of the MNL model to estimate the conditional transition probabilities of crop choice and then generates forecast distributions of total acreages by crop type. For this purpose, we utilize the Cropland Data Layer (CDL), which we overlay with the Common Land Unit (CLU) dataset to accurately define crop choice at the field-level—we focus on the production of corn, soybeans, tobacco, wheat, and alfalfa in Kentucky. 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. Specifically, rotation between soybeans and corn can be explained by changes in prices, but Moreover, in part, by soil productivity. Since corn requires a significant amount of nitrogen fertilizer, farmers tend to incorporate leguminous crops such as soybean into a rotational sequence, which can result in wider forecast distributions. Regarding the forecast distributions for the other crops, they are narrower due
to the fact that alfalfa is a perennial crop and that tobacco and, frequently, wheat are contracted crops. These estimated forecast distributions can be used and applied in various fields of applied research.

The method proposed in this study could be used to evaluate where the most likely places are for the production of new crops such as hemp, inform investments in location specific-capital such as new distilleries, provide information for the management of location-specific input storage distribution such as fertilizer, and assess the likelihood of adverse water quality events such as those triggered by nutrient run-off from crops with significant nutrient applications. Our results may also supplement NASS June Acreage Survey estimates in areas where response rates are low and can serve as interim estimates in the Winter season before the Spring Survey. It is difficult to make direct comparisons between our estimates and those of NASS since NASS combines estimates for several counties Kentucky (i.e., there are only estimates for 79 of 120 counties, and these do not match with counties for which we generate estimates). Despite differences in exact coverage, historical distributions of corn and soybean from 1990 to 2015 based on data from USDA NASS (see Appendix B, Figure A3) show that the forecasted distribution of corn looks similar to the historical distribution of corn even though the average acreage is lower in the forecasted distribution. This can be explained by missing counties in the CLU (See Figure A2). For soybeans, we find that the historical distribution of soybeans is right-skewed, whereas the forecasted distribution of soybeans is normal. Finally, the generation of the forecast distributions is another way for farmers, policymakers, and other stakeholders to consider or visualize uncertainty in forecast estimates by crop.

Several limitations should be outlined alongside our findings. First, 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 is 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 of the CDL data will result in better prediction. Second, 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 the reality of farm production, however, a farmer might grow more than one crop in their field, which needs to be addressed in future studies. Third, our MNL model and simulation exercise are only based on agronomic characteristics. Therefore, this study can be extended by incorporating microeconomic variables such as expected net return, expected price, and farmer’s characteristics. A caveat for this study is that we focus on Kentucky, where agricultural lands are relatively heterogeneous compared to some states like Iowa, Illinois, and Nebraska, where agricultural lands are homogeneous—our results do not easily generalize to other states.

Despite these limitations, our use of the CDL to tackle the specific empirical objective of forecasting agricultural land-use distributions is, to our knowledge, novel and presents several new avenues of applied research.

**Author Contributions:** Conceptualization, G.K., M.N., S.B. and T.M.; methodology, G.K., M.N. and S.B.; software, G.K. and N.M.; validation, G.K., M.N., N.P. and S.B.; formal analysis, G.K. and M.N.; data curation, G.K.; writing—original draft preparation, G.K.; writing—review and editing, G.K., M.N., S.B., N.P. and T.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by National Science Foundation (NSF), grant number 1355438.

**Conflicts of Interest:** The authors declare no conflict of interest.

**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 A2). 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 the merge with the CDL. Second, we had an issue in identifying the crop choice. In realistic, each field is not fully covered by one single crop based on the pixel in CDL when we overlay with CLU (see Figure A1). For instance, there might exist multiple crops in one single field, so we make a strong assumption that the filed is corn if corn acreage dominates other crops. Alternatively, an analyst is able to identify the representative crop in the field by using a centroid point if the point interacts spatially with the pixel. However, this may work if we consider with a large sample such as all 48 contiguous U.S. However, this may provide inaccurate results in forecasting perspectives. Third, some data sets (especially Soil, Slope, and Elevation data) used in this study are not time-varying data; in other words, those data sets are data for a given year. 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.

Appendix B. Additional Figures

Figure A1. 2015 major crops produced in Kentucky from Cropland Data Layer (CDL). Notes: This map is in the field level and created using CDL dataset and filed boundaries. In 2015, Corn was a major crop in Kentucky, followed by Soybeans, Tobacco, Alfalfa, and wheat.
Figure A2. Numbers of missing counties in Kentucky based on CLU.

| 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   | 133  |
| 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   | 238  |

Figure A3. Cont.
Figure A3. Historical distributions of corn and soybeans from 1990 to 2015.

References

1. Hayes, M.; Decker, W. Using NOAA AVHRR data to estimate maize production in the United States Corn Belt. Remote Sens. 1996, 17, 3189–3200. [CrossRef]
2. De Barros Dias, F. US Yield Forecasting Using Crop Condition Ratings. 2017. Available online: https://www.ideals.illinois.edu/bitstream/handle/2142/97265/DEBARROSDIAS-THESIS-2017.pdf?sequence=1&isAllowed=y (accessed on 15 January 2018).
3. Vogel, F.A.; Bange, G.A. Understanding USDA Crop Forecasts. 1999. Available online: https://www.nass.usda.gov/Education_and_Outreach/Understanding_Statistics/pub1554.pdf (accessed on 1 March 2017).
4. Good, D.L.; Irwin, S.H. USDA Corn and Soybean Acreage Estimates and Yield Forecasts: Dispelling Myths and Misunderstandings. 2011. Available online: https://farmdoc.illinois.edu/assets/marketing/mobr/mobr_11-02.pdf (accessed on 17 May 2017).
5. Egelkraut, T.M.; Garcia, P.; Irwin, S.H.; Good, D.L. An evaluation of crop forecast accuracy for corn and soybeans: USDA and private information agencies. J. Agric. Appl. Econ. 2003, 35, 79–95. [CrossRef]
6. Adjemian, M.K.; Smith, A. Using USDA forecasts to estimate the price flexibility of demand for agricultural commodities. Am. J. Agric. Econ. 2012, 94, 978–995. [CrossRef]
7. Good, D.; Irwin, S. Progression of USDA Corn and Soybean Acreage Estimates and Prospects for Final Estimates for 2015. Farmdoc Dly. 2015, 5, 191. Available online: http://farmdocdaily.illinois.edu/2015/10/progression-usda-corn-and-soybean-acreage-estimates.html (accessed on 13 February 2017).
8. Laca, A.-L. Analysts Say the 2019 June Acreage Report Probably Won’t Matter Much. Available online: https://www.agprofessional.com/article/anaylsists-say-2019-june-acreage-report-probably-wont-matter-much (accessed on 20 July 2019).
9. Hendricks, N.P.; Smith, A.; Sumner, D.A. Crop supply dynamics and the illusion of partial adjustment. Am. J. Agric. Econ. 2014, 96, 1469–1491. [CrossRef]
10. Schnepp, R. NASS and US Crop Production Forecasts: Methods and Issues. Congressional Research Service (CRS) Report. 2017. Available online: https://fas.org/sgp/crs/misc/R44814.pdf (accessed on 20 April 2018).
11. Johansson, R.; Effland, A.; Coble, K. Falling response rates to USDA crop surveys: Why it matters. Farmdoc Dly. 2017, 7, 1–6.
12. Lark, T.; Hendricks, N.P.; Pates, N.; Smith, A.; Spawn, S.; Bougie, M.; Booth, E.; Kucharik, C.; Gibbs, H. Impacts of the Renewable fuel Standard on America’s Land and Water Resources. 2019. Available online: https://aaas.confex.com/aaas/2019/poster/eposter.cgi?eposterid=549 (accessed on 20 July 2019).
13. Martinez, A.; Maier, D.E. Quantification of biomass feedstock availability to a biorefinery based on multi-crop rotation cropping systems using a GIS-based method. *Biol. Eng. Trans.* 2014, 7, 3–16.

14. Leteinturier, B.; Herman, J.-L.; De Longueville, F.; Quintin, L.; Oger, R. Adaptation of a crop sequence indicator based on a land parcel management system. *Agric. Ecosyst. Environ.* 2006, 112, 324–334. [CrossRef]

15. Porter, P.M.; Lauer, J.G.; Luoschen, W.E.; Ford, J.H.; Hoverstad, T.R.; Oplinger, E.S.; Crookston, R.K. Environment affects the corn and soybean rotation effect. *Agron. J.* 1997, 89, 442–448. [CrossRef]

16. Hendricks, N.P.; Sinnathamby, S.; Douglas-Mankin, K.; Smith, A.; Sumner, D.A.; Earnhart, D.H. The environmental effects of crop price increases: Nitrogen losses in the US Corn Belt. *J. Environ. Econ. Manag.* 2014, 68, 507–526. [CrossRef]

17. Karlen, D.L.; Hurley, E.G.; Andrews, S.S.; Cambardella, C.A.; Meek, D.W.; Duffy, M.D.; Mallarino, A.P. Crop rotation effects on soil quality at three northern corn/soybean belt locations. *Agron. J.* 2006, 98, 484–495. [CrossRef]

18. Plourde, J.D.; Pijanowski, B.C.; Pekin, B.K. Evidence for increased monoculture cropping in the Central United States. *Agric. Ecosyst. Environ.* 2013, 165, 50–59. [CrossRef]

19. Halvorson, A.D.; Del Grosso, S.J.; Reule, C.A. Nitrogen, tillage, and crop rotation effects on nitrous oxide emissions from irrigated cropping systems. *J. Environ. Qual.* 2008, 37, 1337–1344. [CrossRef] [PubMed]

20. Burgess, M.H.; Miller, P.R.; Jones, C.A. Pulse crops improve energy intensity and productivity of cereal production in Montana, USA. *J. Sustain. Agric.* 2012, 36, 699–718. [CrossRef]

21. Long, J.A.; Lawrence, R.L.; Miller, P.R.; Marshall, L.A. Changes in field-level cropping sequences: Indicators of shifting agricultural practices. *Agric. Ecosyst. Environ.* 2014, 189, 11–20. [CrossRef]

22. McFadden, D. Conditional Logit Analysis of Qualitative Choice Behavior. In *Frontiers in Econometrics*; Academic Press: New York, NY, USA, 1973; pp. 105–142.

23. Carrión-Flores, C.E.; Flores-Lagunes, A.; Guici, L. Land Use Change: A Spatial Multinomial Choice Analysis. In Proceedings of the 2009 Annual Meeting, Milwaukee, WI, USA, 26–28 July 2009. Available online: https://ageconsearch.umn.edu/record/49403/files/CFG_AAEA_612952.pdf (accessed on 10 January 2010).

24. Hardie, I.W.; Parks, P.J. Land use with heterogeneous land quality: An application of an area base model. *Am. J. Agric. Econ.* 1997, 79, 299–310. [CrossRef]

25. Plantinga, A.J.; Mauldin, T.; Miller, D.J. An econometric analysis of the costs of sequestering carbon in forests. *Am. J. Agric. Econ.* 1999, 81, 812–824. [CrossRef]

26. Wu, J.; Adams, R.M.; Kling, C.L.; Tanaka, K. From microlevel decisions to landscape changes: An assessment of agricultural conservation policies. *Am. J. Agric. Econ.* 2004, 86, 26–41. [CrossRef]

27. Paton, L.; Troffaes, M.C.; Boatman, N.; Hussein, M.; Hart, A. 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: Cham, UK, 2014; pp. 476–485.

28. Matis, J.; Saito, T.; Grant, W.; Iwig, W.; Ritchie, J. A Markov chain approach to crop yield forecasting. *Agric. Syst.* 1985, 18, 171–187. [CrossRef]

29. Boryan, C.; Yang, Z.; Mueller, R.; Craig, M. Monitoring US agriculture: The US department of agriculture, national agricultural statistics service cropland data layer program. *Geocarto Int.* 2011, 26, 341–358. [CrossRef]

30. Stern, A.; Doraiswamy, P.C.; Hunt, E.R. Changes of crop rotation in Iowa determined from the United States Department of Agriculture, National Agricultural Statistics Service cropland data layer product. *J. Appl. Remote Sens.* 2012, 6, 1–16. [CrossRef]

31. Yost, M.A.; Russell, M.P.; Coulter, J.A.; Bolstad, P.V. Alfalfa stand length and subsequent crop patterns in the upper midwestern United States. *Agron. J.* 2014, 106, 1697–1708. [CrossRef]

32. Lusk, J.L.; Tack, J.; Hendricks, N.P. Heterogeneous Yield Impacts from Adoption of Genetically Engineered Corn and the Importance of Controlling for Weather; 0898-2937; National Bureau of Economic Research: Cambridge, MA, USA, 2019; pp. 11–39.

33. Croissant, Y. Estimation of multinomial logit models in R: The mlogit Packages. R package version 0.2-2. 2012. Available online: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.303.7401&rep=rep1&type=pdf (accessed on 20 March 2017).

34. Bell, E.J. Markov analysis of land use change—an application of stochastic processes to remotely sensed data. *Socio-Econ. Plan. Sci.* 1974, 8, 311–316. [CrossRef]

35. Baker, W.L. A review of models of landscape change. *Landsc. Ecol.* 1989, 2, 111–133. [CrossRef]
36. Brown, D.G.; Pijanowski, B.C.; Duh, J.D. Modeling the relationships between land use and land cover on private lands in the Upper Midwest, USA. *J. Environ. Manag.* 2000, 59, 247–263. [CrossRef]

37. Muller, M.R.; Middleton, J. A Markov model of land-use change dynamics in the Niagara Region, Ontario, Canada. *Landscape Ecol.* 1994, 9, 151–157.

38. Savage, T. Forecasting Land Use from Estimated Markov Transitions. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1866003 (accessed on 14 March 2017).

39. Thornton, P.K.; Jones, P.G. A conceptual approach to dynamic agricultural land-use modelling. *Agric. Syst.* 1998, 57, 505–521. [CrossRef]

40. Aurbacher, J.; Dabbert, S. Generating crop sequences in land-use models using maximum entropy and Markov chains. *Agric. Syst.* 2011, 104, 470–479. [CrossRef]

41. Castellazzi, M.; Wood, G.; Burgess, P.J.; Morris, J.; Conrad, K.; Perry, J. A systematic representation of crop rotations. *Agric. Syst.* 2008, 97, 26–33. [CrossRef]

42. Osman, J.; Inglada, J.; Dejoux, J.-F. Assessment of a Markov logic model of crop rotations for early crop mapping. *Comput. Electron. Agric.* 2015, 113, 234–243. [CrossRef]

43. Troffaes, M.; Paton, L. *Logistic Regression on Markov Chains for Crop Rotation Modelling*; Society for Imprecise Probability: Theories and Applications (SIPTA): Compiegne, France, 2013; pp. 329–336.

44. Taylor, H.; Karlin, S. *An Introduction to Stochastic Modeling*; Academic Press: New York, NY, USA, 1984.

45. Louviere, J.J.; Hensher, D.A.; Swait, J.D. *Stated Choice Methods: Analysis and Applications*; Cambridge University Press: Cambridge, UK, 2000.

46. Domenech, T.; McFadden, D. *Urban Travel Demand: A Behavioral Analysis, with T; Domencich*. North Holland: Amsterdam, The Netherlands, 1975.

47. The Fertilizer Institute. 2018 State of the Fertilizer Industry. Available online: https://www.tfi.org/sites/default/files/tfi-stateoftheindustry-2018.pdf (accessed on 20 July 2019).

48. Coomes, P.; Kornstein, B. The Economic and Fiscal Impacts of the Distilling Industry in Kentucky. Available online: https://kybourbon.com/wp-content/uploads/2019/02/2019-Economic-Impact-Report_FINAL.pdf (accessed on 20 July 2019).