Advances in land-use and stated-choice modeling using neural networks and discrete-choice models

by

Steven M. Ramsey

B.S. Economics, University of Kansas, 2007

M.A. Economics, University of Kansas, 2009

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Agricultural Economics
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2017

Abstract

Applied research in agricultural economics often involves a discrete process. Most commonly, these applications entail a conceptual framework, such as random utility, that describes a discrete-variable data-generating process. Assumptions in the conceptual framework then imply a particular empirical model. Common approaches include the binary logit and probit models and the multinomial logit when more than two outcomes are possible. Conceptual frameworks based on a discrete choice process have also been used even when the dependent variable of interest is continuous. In any case, the standard models may not be well suited to the problem at hand, as a result of either the assumptions they require or the assumptions they impose. The general theme of this dissertation is to adopt seldom-used empirical models to standard research areas in the field through applied studies. A common motivation in each paper is to lessen the exposure to specification concerns associated with more traditional models.

The first paper is an attempt to provide insights into what — if any — weather patterns farmers respond to with respect to cropping decisions. The study region is a subset of 11 north-central Kansas counties. Empirically, this study adopts a dynamic multinomial logit with random effects approach, which may be the first use of this model with respect to farmer land-use decisions. Results suggest that field-level land-use decisions are significantly influenced by past weather, at least up to ten years. Results also suggest, however, that that short-term deviations from the longer trend can also influence land-use decisions.

The second paper proposes multiple-output artificial neural networks (ANNs) as an alternative to more traditional approaches to estimating a system of acreage-share equations. To assess their viability as an alternative to traditional estimation, ANN results are compared to a linear-in-explanatory variables and parameters heteroskedastic and time-wise autoregressive seemingly unrelated regression model. Specifically, the two approaches are

compared with respect to model fit and acre elasticities. Results suggest that the ANN is a viable alternative to a simple traditional model that is misspecified, as it produced plausible acre-response elasticities and outperformed the traditional model in terms of model fit.

The third paper proposes ANNs as an alternative to the traditional logit model for contingent valuation analysis. With the correct network specifications, ANNs can be viewed as a traditional logistic regression where the index function has been replaced by a flexible functional form. The paper presents methods for obtaining marginal effect and willingness-to-pay (WTP) measures from ANNs, which has not been provided by the existing literature. To assess the viability of this approach, it is compared with the traditional logit and probit models as well an additional semi-nonparametric estimator with respect to model fit, marginal effects, and WTP estimates. Results suggest ANNs are viable alternative and may be preferable if misspecification of the index function is a concern.

Advances in land-use and stated-choice modeling using neural networks and discrete-choice models

by

Steven M. Ramsey

B.S. Economics, University of Kansas, 2007

M.A. Economics, University of Kansas, 2009

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Agricultural Economics
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2017

Approved by:
Co-Major Professor
Jason S. Bergtold

Approved by:
Co-Major Professor
Jessica Heier Stamm

Abstract

Applied research in agricultural economics often involves a discrete process. Most commonly, these applications entail a conceptual framework, such as random utility, that describes a discrete-variable data-generating process. Assumptions in the conceptual framework then imply a particular empirical model. Common approaches include the binary logit and probit models and the multinomial logit when more than two outcomes are possible. Conceptual frameworks based on a discrete choice process have also been used even when the dependent variable of interest is continuous. In any case, the standard models may not be well suited to the problem at hand, as a result of either the assumptions they require or the assumptions they impose. The general theme of this dissertation is to adopt seldom-used empirical models to standard research areas in the field through applied studies. A common motivation in each paper is to lessen the exposure to specification concerns associated with more traditional models.

The first paper is an attempt to provide insights into what — if any — weather patterns farmers respond to with respect to cropping decisions. The study region is a subset of 11 north-central Kansas counties. Empirically, this study adopts a dynamic multinomial logit with random effects approach, which may be the first use of this model with respect to farmer land-use decisions. Results suggest that field-level land-use decisions are significantly influenced by past weather, at least up to ten years. Results also suggest, however, that that short-term deviations from the longer trend can also influence land-use decisions.

The second paper proposes multiple-output artificial neural networks (ANNs) as an alternative to more traditional approaches to estimating a system of acreage-share equations. To assess their viability as an alternative to traditional estimation, ANN results are compared to a linear-in-explanatory variables and parameters heteroskedastic and time-wise autoregressive seemingly unrelated regression model. Specifically, the two approaches are

compared with respect to model fit and acre elasticities. Results suggest that the ANN is a viable alternative to a simple traditional model that is misspecified, as it produced plausible acre-response elasticities and outperformed the traditional model in terms of model fit.

The third paper proposes ANNs as an alternative to the traditional logit model for contingent valuation analysis. With the correct network specifications, ANNs can be viewed as a traditional logistic regression where the index function has been replaced by a flexible functional form. The paper presents methods for obtaining marginal effect and willingness-to-pay (WTP) measures from ANNs, which has not been provided by the existing literature. To assess the viability of this approach, it is compared with the traditional logit and probit models as well an additional semi-nonparametric estimator with respect to model fit, marginal effects, and WTP estimates. Results suggest ANNs are viable alternative and may be preferable if misspecification of the index function is a concern.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to first thank my major professors, Dr. Jason Bergtold and Dr. Jessica Heier-Stamm, for the examples they have provided as researchers; mentors and teachers; professionals; and friends. Working closely with them over the years, I have learned much and seen much to admire that goes beyond just the technical details of an education. I would also like to thank the other members of my committee, Dr. Marcellus Caldas and Dr. Nelson Villoria for their help and support. I am also thankful to Dr. Christine Wilson for her help and encouragement with my teaching development; and to the department for offering me these opportunities. Thank you Dr. Nathan Hendricks and Dr. Jesse Tack, for the friendships and willingness to provide advice and insights, as well as the insightful comments and questions. I would also like to thank Dr. Rich Llewelyn for his assistance in providing and helping with data used in this study. Thank you to Deana Foster, Judy Maberry, Carla Woodward, Terry Teske and the rest of the staff who helped to make my life easier.

Thanks to all of the fellow graduate students, past and present, who have made my time here at Kansas State enjoyable and memorable. A special thank you to Melissa McKendree, Ana Claudia Sant'Anna, and Frank Nti; who have helped me and endured with me through the major hurdles of the program. Also to Nicholas Pates for encouraging me to expand my computing repertoire and for other insights along the way, and also for the occasional trip to Aggieville. My best wishes to you all in your future endeavors.

Finally, I am eternally grateful for the love and support of my family throughout this process. I thank my mom and dad (Jeanne and Ken) for their moral support; financial support; and so many ancillary activities, such as dog-sitting, which have helped me through the process. I thank my brother (Andrew) and my sister (Kendra) and her husband (Kyle) and their families (Layla, Ava, and Phog) for keeping me grounded, offering me perspective, and providing an escape from the daily grind. Thank you also to my grandparents, Keith and Jo Anne McKee and Joe and Connie Ramsey for their love and support. Finally, I thank my best friend of these last years, Kota — the greatest dog that has ever or will ever walk the face of the Earth — for sacrificing games of fetch or tug of war as I put in long hours, for always putting a smile on my face, keeping me sane, and reminding me to be thankful for those things that truly matter.

This research was supported by the CNH: Coupled Climate, Cultivation and Culture in the Great Plains: Understanding Water Supply and Water Quality in a Fragile Landscape, NSF, CNH Award #1313815.

Dedication

To my family.

Chapter 1

Introduction

Applied research in agricultural economics often involves a discrete choice process. Commonly, these applications entail a conceptual framework, such as random utility, that describes a discrete data-generating process. Assumptions in the conceptual framework then imply a particular empirical model. For binary outcomes, the implied empirical model is typically the standard logit or probit type model. A typical application for these models may be the examination of stated choice survey data wherein respondents have "accepted" or "rejected" the option to procure a good or government policy. The estimated models can then be used for predictions, marginal effects on policy support, willingness-to-pay for a good or policy action, etc. When the discrete outcome has more than two possibilities, the most widely used model is the multinomial logit due to estimation difficulties associated with the multinomial probit (W. H. Greene, 2012). Multinomial models may be used for analyses similar to those mentioned for the binary versions and are also commonly used in land-use decision or land-use transition studies.

Even when the dependent variable is not discrete, a conceptual framework based on a discrete choice process can still be used. A prominent example of this is the linear-logit model and its subsequent extension by Theil (1969) to the case of more than two outcomes. Analyses employing this approach are often interested in modeling variable shares, e.g., expenditure shares on specific categories or land-use shares within a region. To this end, researchers assume the observed share is equal to the probability of some specific event, say the probability a generic field within the region of interest being allocated to wheat. This assumption can allow the standard multinomial logit framework to be used as a starting point. Then, following a couple of simple transformations, the researcher is able to examine what drives particular shares using standard regression techniques.

Often, however, the standard models are not well suited to the problem at hand as a result of either the assumptions they require or the assumptions they impose. With respect to the multinomial logit model, the imposition of the independence from irrelevant alternatives property is commonly noted as a drawback. This property, which assumes the relative odds of choosing one option over another are independent from other alternatives and from the attributes associated with those alternatives, is not appropriate in many situations (Train, 2003). As a result, alternative specifications have been offered to relax this assumption, such as the nested, latent class, and random parameter logit models (Wooldridge, 2010).

Additional complications can arise through the use of panel data, which may result in a violation of the model's underlying assumptions. Consider as an example the binary case where there is time-invariant unobserved heterogeneity across cross-section units. Under this scenario, pooled estimation will yield parameter estimates that are inconsistent (W. H. Greene, 2012). The presence of unobserved heterogeneity can also create an endogenous variable problem when lagged-dependent variables are included on the right-hand side of a model. When unobserved heterogeneity is assumed, the inclusion of lagged effects can present what is known as the initial conditions problem. With relatively short panels, the initial conditions — i.e., the earliest observation on the dependent variables — can play a crucial role in the outcomes path (W. H. Greene, 2012). Thus, the standard estimators, as well as many of their extensions, will not be consistent (W. H. Greene, 2012).

Heckman (1981) observed that initial conditions were typically treated by assuming either that they are truly exogenous variables or that the stochastic process is in equilibrium. In most applications, it is unlikely that either assumption holds. With land-use decisions, for example, the first assumption is impractical as it assumes disturbances in the model are serially independent (Heckman, 1981). Given the importance of possible temporal (rotation) effects and also the possibility of unobserved heterogeneity (e.g., across farmers or fields), this assumption is unrealistic. As to the second assumption, Heckman (1981) notes it is unlikely to hold when the process is driven by time-varying exogenous variables. Continuing with the land-use decision model, this too is unlikely when we consider prices, rotations (the lagged-dependent variables), weather, etc. In response to these unsatisfying solutions, Heckman (1981) proposed an approach which relies upon approximating the conditional distribution of the initial conditions (Wooldridge, 2005).

Wooldridge (2005), however, claimed that obtaining parameter estimates and average effects are more computationally burdensome than need be under the Heckman (1981) approach. Instead, Wooldridge (2005) proposed using the distribution of the unobserved effect given the initial condition and any exogenous variables. With this framework, Wooldridge (2005) dealt with the initial conditions problem with a standard random-effects binary-probit model. Subsequent studies have adapted this approach to multiple-outcome scenarios through the use of a "dynamic-multinomial logit"[1], e.g., Bjørner and Leth-Petersen (2007); Chatterjee (2011); and Fok, Jeon, and Wilkins (2013). To date, it appears that neither the approach from Heckman (1981) nor that from Wooldridge (2005) has been used in an analysis of land-use decisions with multiple potential outcomes.

Another concern for studies built around a discrete choice framework (and any other econometric model) is misspecification. Within the discrete choice framework, distributional assumptions, e.g., opting for the logit model over the probit model, are one cause for misspecification concerns. In terms of model results with a binary outcome, e.g., predictions or marginal effects, the consequences of choosing the logit when the true model is the probit are relatively benign, except when data are heavily concentrated in the tails (Amemiya, 1981). With multiple outcome possibilities, however, the logit/probit decision can result in more substantial differences (Amemiya, 1981), though often the multinomial logit is used due to the previously mentioned estimation challenges presented by the multinomial probit.

From a theoretical perspective, model choice — even in the binary case — can be much

[1] A similar approach of the same name was presented by Gong, Van Soest, and Villagomez (2004) and is based on the approach from Heckman (1981) (W. Greene, 2012).

more important. For the univariate-binary logit model, Kay and Little (1987) provide explanatory-variable transformations that are necessary to be consistent with a logit model which is linear in both parameters and the explanatory variables. The transformations from Kay and Little (1987) are dependent on the the distribution of the explanatory variables conditioned on the dependent variable. With respect to the probit model, no such results are available. Thus, the statistical consistency of an index function that is linear in parameters and explanatory variables — or any other specification — is difficult to discern (Bergtold, Spanos, & Onukwugha, 2010). Even if the researcher has made the correct distributional assumption, the results from Kay and Little (1987) and extensions from Bergtold et al. (2010) suggest that misspecification is still likely present in the index function. Though an incorrect distributional assumption in the linear-logit models is a moot point — the model does not exist if this assumption does not hold — misspecification of the index function is still relevant.

Misspecification resulting from either the distributional or index function assumptions are clearly related concepts. If a researcher is concerned with either (or both), semi-nonparametric (SNP) and non-parametric (NP) alternatives are available, such as the SNP estimators of Gallant and Nychka (1987); Klein and Spady (1993); and Creel and Loomis (1997) for binary outcomes. SNP approaches have the benefit of being less restrictive than fully-parametric models, such as the logit or probit, and may avoid some of the drawbacks of NP models such as decreased estimation precision, results which are difficult to display or communicate, and the inability to extrapolate or impose restrictions suggested by economic theory (Horowitz, 2009). However, use of many SNP estimators may be limited as they are not readily available in common statistical software packages. Additionally, for many SNP approaches, it is not possible to estimate average partial effects and discrete explanatory variables are not permitted (Wooldridge, 2010).

Misspecification of the index function, with regards to either the linear-logit for continuous outcomes or traditional logit for binary outcomes, can be addressed through another less commonly used SNP approach, artificial neural networks (ANNs). The basic idea is that a specific, assumed functional form, e.g., $\mathbf{x}'\boldsymbol{\beta}$, is replaced with a flexible functional form. Due to their capabilities in classification and function approximation, ANNs present an attractive alternative in both cases. Moreover, marginal effects (or average partial effects) are attainable analytically via the chain rule of differentiation. ANNs are also increasingly available as built-in routines in many software packages, such as MATLAB, R, and SAS. Thus, there is potential for their use across a broad range of applied studies.

The general theme of this dissertation is to adopt these seldom-used (at least within the current contexts) empirical models — the dynamic-multinomial logit based on Wooldridge (2005) and ANNs — to common research areas in the fields of land-use decisions and stated choice analysis. The proposed empirical methods provide researchers with readily available alternative approaches that can reduce misspecification risk in many common applications. Thus, the offered approaches can help to avoid potentially inaccurate or inconsistent estimates and inferences. Specifically, the objective of this dissertation is to apply a dynamic-multinomial logit with random effects to field-level land-use decisions and to present ANNs as alternatives to (1) the linear-logit model for acreage-share analysis and (2) the traditional-binary logit model for a stated-choice analysis. A brief description of each paper is provided below.

1.1 Paper 1: Field-level land-use adaptation to local weather trends

The intersection of agriculture and climate has been well researched for at least the last couple of decades. Largely, the motivation for previous research has been the potential impacts on food security for the world's (growing) population. Many studies have predicted unfavorable yield scenarios for particular geographic regions. As a result, another common research theme is farmer adaptation to a changing climate. Typically, these studies are concerned with what farmers *could* or *should* do to adapt to adverse outcomes. However, research examining whether or not farmers actually do respond to weather patterns has largely been ignored. Answers to this question can help provide more accurate food security analysis: if farmers *do* respond to changing patterns through say cropping decisions, the global food supply outcome will be different than a world in which they *do not* respond.

This study is an attempt to provide insights into what — if any — weather patterns farmers respond to with respect to cropping decisions. The study region is small — a set of 11 Kansas counties — and so extrapolation of results to larger scales may not be warranted. However, the study is an important step towards more credible estimates of global food supplies under changing climates and the methods themselves translate to other areas. Results of the study may also be of interest to community members of the 11 counties, the Kansas community at large, Kansas policy makers, or anyone generally who is interested in agriculture/climate dynamics. Empirically, this study adopts a dynamic-multinomial logit approach based on the work of Wooldridge (2005) that — to the author's knowledge — is the first use of this model with respect to farmer land-use decisions.

1.2 Paper 2: An artificial neural network approach to acreage-share modeling

Agricultural land-use patterns impact societies in many ways. For example, micro- or macro-level supply or demand shocks that, in turn, impact individual livelihoods and global food security. Agricultural land-use can also impact local or global environments. Often, empirical analyses will examine the factors that drive the shares of particular land uses within predefined geographic boundaries, such as states or counties. Because shares are bounded between zero and one, a natural approach to modeling shares is through a cumulative density function (cdf). By choosing the logistic cdf, researchers are able to use simplified regression techniques through a series of transformations that result in what is known as the linear-logit model. Estimation then typically proceeds by assuming a linear index function associated with the cdf and finally estimating the model via seemingly unrelated regression (SUR) to account for contemporaneous correlation. An extension to this approach was provided by Wu and Brorsen (1995): the heteroskedastic and time-wise autoregressive SUR (SUR-

HEAR). The motivation behind the SUR-HEAR was to account for autocorrelation and heteroskedasticity in addition to contemporaneous correlation.

The primary objective of this paper is to propose the use of multiple-output artificial neural network (ANNs) as an alternative to more traditional approaches (i.e., SUR and SUR-HEAR) to estimating a system of acreage-share equations. Commonly, ANNs are used in studies primarily concerned with prediction or forecasting. The use of ANNs towards these ends arises from their performance, which can be attributed to their function approximation and classification capabilities. For this study, the individual outputs from the ANN serve as flexible functional form approximations of the true underlying index function. To assess their viability as an alternative to SUR-HEAR estimation of the linear logit, ANN results are compared to a SUR-HEAR model which assumes a linear-in-explanatory variables and linear-in-parameters index function. Specifically, the two approaches are compared with respect to model fit and crop-acreage elasticities.

1.3 Paper 3: Neural network estimators of binary choice processes for valuing environmental amenities: Estimation, marginal effects, and WTP

In the field of environmental economics, stated choice surveys are often administered with the goal of measuring community support for some policy that will protect or enhance an environmental service or amenity. In the binary case, where respondents are faced with two options, "accept" or "reject," the most common modeling approaches are to use either the binary logit or binary probit models. Obtaining estimates of marginal effects or willingness-to-pay (WTP) for the proposal are often of equal or greater importance than predicting responses. Whatever the underlying motivations, results from a statistically misspecified model can lead to inaccurate inferences.

This paper proposes artificial neural networks (ANNs) as a semi-nonparametric (SNP) alternative to the traditional logit model. With the correct network specifications, ANNs can be viewed as a traditional logistic regression where the index function has been replaced by a flexible functional form. In other words, this paper offers a method for avoiding misspecification with respect to the index function. Additionally, the paper presents methods for obtaining marginal effects and WTP measures from ANNs, which — to the author's knowledge — has not been provided by the existing literature. The ability to generate marginal effects and WTP measures from ANNs, combined with their increasing availability in many statistical software packages, is an advantage over other SNP binary choice models, which either may not provide marginal effects, may not be widely accessible, or both. To assess the viability of this approach, it is compared with the traditional logit and probit models as well as the SNP estimator from Klein and Spady (1993) with respect to model fit, marginal effects, and WTP estimates.

Chapter 2

Field-level land-use adaptation to local weather trends

2.1 Introduction

It is important to be aware of potential climate-change implications, regardless of the beliefs one holds about the validity of recent climate change or projected climate-change scenarios. This is certainly true with respect to potential implications for agriculture, at least so long as we are concerned with food security or families and communities who are dependent on agriculture. Given the importance of agriculture, combined with its susceptibility to adverse — including increasingly variable — weather patterns, it is not surprising that the intersection of agriculture and climate has been well researched; see Mendelsohn, Nordhaus, and Shaw (1994); Lal (2004); Long, Ainsworth, Leakey, Nösberger, and Ort (2006); Lobell et al. (2008); Searchinger et al. (2008); Schlenker and Roberts (2009), to name just a few of the notable examples.

Much research to date has focused on yield impacts on food supplies, e.g., Rosenzweig, Parry, et al. (1994); Long et al. (2006); Lobell et al. (2008); and Schlenker and Roberts (2009). Though many of these studies do not paint an overall-favorable picture for global or regional food supply under the assessed climate scenarios, impacts tend to be spatially heterogeneous, with "winners" and "losers" determined by local phenomena. There has also been considerable research regarding potential adaptation and/or mitigation strategies for agriculture in the face of climate change, such as Smit and Skinner (2002); Bradshaw, Dolan, and Smit (2004); Howden et al. (2007); and B. B. Lin (2011). Still others have examined climate change perceptions and beliefs held by farmers and the factors that shape them, e.g., Diggs (1991); Haden, Niles, Lubell, Perlman, and Jackson (2012); and Arbuckle Jr, Morton, and Hobbs (2015).

With respect to food security, impacts on crop yields are only part of the story, as total supplies are determined by yields and planted acreage. In light of this, what appears to be lacking in the current literature are empirical analyses investigating whether farmer land-use decisions *have* responded to changes in weather patterns, rather than if they *should* or *could* do so. The relevance of whether farmers recognize changing climates and respond accordingly is compounded as most countries prefer an incentive-based rather than regulation-based approach to adaptation or mitigation (Haden et al., 2012). This paper investigates whether land-use decisions in a subset of Kansas counties have been influenced by precipitation trends and, if so, what time horizons are the most important. Through the inclusion of different precipitation histories, insights are gained as to how quickly (or slowly) agricultural production in a region may respond or adapt to a changing climate.

2.2 Background

Common motivations for previous research on the agriculture-climate change relationship include food security concerns and agriculture's potential to compound or mitigate adverse climate change scenarios. With respect to the first motivation, examples in the literature are plenty. Long et al. (2006) review free-air concentration enrichment technology studies to challenge previous findings and conclude that global yield decreases for crops including corn, rice, sorghum, soybeans, and wheat from increased temperatures and decreased soil moisture are not likely to be offset by yield increases due to direct fertilization from rising carbon-dioxide (CO_2) levels. With growing regions fixed, Schlenker and Roberts (2009) predict global-yield decreases by 2099 of between 30 and 82% for corn, soybeans, and cotton, depending on the speed at which climates change. Lobell, Schlenker, and Costa-Roberts (2011) find that from 1980 to 2008, relative to a no-climate-trend counter-factual, global maize and wheat production declined by 3.8 and 5.5% respectively, while soybeans and rice were relatively unchanged. Parry, Rosenzweig, Iglesias, Livermore, and Fischer (n.d.) use climate change projections and yield functions for maize, rice, soybeans, and wheat to estimate the impact of climate scenarios on the number of people at risk from hunger. Though the authors find global production to be relatively stable, the estimated distribution of impacts — with negative impacts skewed towards developing countries — leads to substantial increases in prices and risk of hunger for poorer nations.

Searchinger et al. (2008) provides one example where agriculture is identified as a contributor to climate change. In this study, the authors conclude that biofuel mandates have increased greenhouse-gas (GHG) emissions through the induced conversion of forests and grasslands to crop production. Though the drivers in this case are (perhaps misguided) governmental policies, a connection between agriculture and climate change through unintended consequences was created. Additional impacts from agriculture are given by Bellarby, Foereid, and Hastings (2008), such as GHG emissions from tillage, chemical applications, drilling or seeding, harvesting, and irrigation. Bellarby et al. (2008) also note indirect GHG impacts from related industries such as pesticide and fertilizer production.

Due to potential food-security ramifications and, to a lesser extent, the role agriculture may play in compounding adverse scenarios, numerous studies have looked at climate-change adaptation or mitigation strategies that could be employed at the farm level. The focus of this study is adaptation, which may be realized in numerous ways and, to quote Smit and Skinner (2002, p. 86), may "encompass a wide range of forms (technical, financial, managerial), scales (global, regional, local) and participants (governments, industries, farmers)." Olesen and Bindi (2002, p. 252-253) identify several adaptation strategies including "short-term adjustments" such as modified planting schedules, input adjustments, and water-conserving practices such as reduced-tillage or irrigation management; and "long-term adaptations" such as changes in land allocations or a move away from specialized production. With respect to changes in land allocations, Olesen and Bindi (2002) suggest that crops with highly variable production could be substituted for crops with more stable yields, even if total production is lower. This study examines mitigation only to the extent that land-use decisions are considered a mitigation strategy. Smith et al. (2008) provide a few examples where this can hold, such as avoiding fallow periods; extending rotation periods, particularly those with perennial crops; or the use of rotations with legume crops. Other forms of mitigation include changes in nutrient management, tillage, or water management; and species introduction, just to name a few (Smith et al., 2008). Given the data set used in this study, it is not possible to identify management-based mitigation strategies, such as tillage or species changes. With respect to land-use changes, it seems the more conservative assumption is to attribute weather-induced land-use responses to adaptation rather than mitigation. That is, it seems safer to assume land-use decisions are made from profit-type perspective rather than a GHG-emissions-reduction perspective.

Assuming the area of interest for this study, shown in figure 2.1, is not resilient to climate trends, then agricultural production here will likely be affected by adaptation-motivating forces created by a changing climate. Adams et al. (1990) simulated agricultural impacts based on two different climate models using a doubled CO_2 scenario. Across the two models, the Northern Plains (NP) region — Kansas, Nebraska, North Dakota, and South Dakota — was subject to annual precipitation changes of between -3 and $+7\%$ and average annual temperature increases of 4.7 to 5.9 degrees Celsius. Using the various scenarios, the authors predicted increased crop yields in some cases; but also an increase in irrigated acreage of between 3.94 and 4.03 million acres. While not a "climate change" study, per se, Staggenborg, Dhuyvetter, and Gordon (2008) offer additional insights as to what impacts changing weather patterns could have for Kansas farmers. Using field data from research plots and producer fields, Staggenborg et al. (2008) compared corn and sorghum yields to determine if either crop offers an advantage in drought conditions. The authors concluded that sorghum may be a better choice in areas with erratic rainfall and high temperatures. Thus, if Kansas farmers are faced with increased rainfall variability and increased temperatures, they could potentially replace corn acres with sorghum acres. Evidence from Long et al. (2006) meanwhile suggests that soybean and wheat yields may respond more positively under typical climate change scenarios relative to corn and sorghum. Drought stress and extreme heat stress are likely to become more frequent under many climate change scenarios (M. P. Reynolds et al., 2016), thus, incentives to adapt at the farm level based on annual-crop performance may be more frequent as well.

Despite the potential global ramifications and localized incentives to adapt, farmers may not respond in ways or at a rate the scientific community would prefer. One reason is that most climate-change scenarios or measures are global, and thus may have little actual or perceived relevance to individual farmers (Morton, Hobbs, Arbuckle, & Loy, 2015). Another, somewhat related, reason could be a lack of faith in future projections, which is understandable given the accuracy of short-term forecasts. Weather forecasts 10 days out, for example, may only have a 40% forecast skill, defined as the correlation between forecasts and a verifying analysis (Bauer, Thorpe, & Brunet, 2015). Though the evidence from Bauer et al. (2015) is with respect to tropical zones, there is likely a correlation between predictive accuracy in those zones and predictive accuracy in other areas, e.g., Kansas. The validity of climate change has also increasingly become a point of political division, which has likely exacerbated initial skepticism for some farmers. Whatever the reasons, previous research indicates a sizable portion of the farming community has doubts regarding climate change, which is an obvious barrier to adaptation. Arbuckle, Prokopy, et al. (2013), for example, in a survey of farmers across 11 states found that 34.5% of farmers were uncertain or did not believe climate change is occurring. In a survey of just Iowa farmers, Arbuckle, Morton, and Hobbs (2013) found similar results, with 32.1% of respondents holding a view of uncertainty or disbelief in the notion that climate change is occurring. In both cases, farmers who believed climate change is occurring were more likely to support adaptive or mitigative measures.

Regardless of skepticism around the issue, we should expect that individual farmers may react to local climate change if it results in negative economic outcomes. Thus, even the most ardent skeptic may undertake adaptation measures in an effort to maintain economic viability. The question then becomes: are farmers responding to weather patterns? And, if so, to what are they responding? To put this second question another way, Bradshaw et al.

(2004) ask "will farmers differentiate between the so-called 'signal' and 'noise'?" In other words, will farmers differentiate between long-term weather trends (the signal) and annual variations around this trend (the noise)? Bradshaw et al. (2004) note that the answer to this question may be unimportant if farmers' adaptations leaves them better prepared for long-term trends, but important if farmers respond to variability which masks slower, underlying trends. Bradshaw et al. (2004) suggest the latter case could lead to complacency, inefficient adaptation, or maladaptation.

The question of "to what" — if anything — are farmers responding is the central question of this study. Attempts to answer this question are done primarily through the lens of time, i.e., "If farmers are responding to weather, then what historical time windows are the most important?" An answer to this question will provide insights as to the speed at which farmers (who may be skeptical about climate change) might employ adaptation measures. It may also provide an answer to the question posed by Bradshaw et al. (2004). Additionally, results of this study will provide insights into how land-use patterns may shift under various climate change scenarios. Taking both insights together, results from this study can guide future researchers in conducting more thorough assessments of climate-change impacts.

2.3 Conceptual Framework

The conceptual model starts with farmer k, assumed to be a profit maximizer. Farmer k may operate on more than one field, and each year must decide what crop is to be planted in each field. Assume that a given field i is allocated to only one land use in a given year t. The farmer's problem then is to maximize field-level expected profits across land uses. Letting $j = 0, \ldots, J$ denote the crop choices available, farmer k's problem in year t for field

i can be represented with the following:

$$\begin{aligned}
\max \Pi_t &= \textstyle\sum_j d_{i,t}^j \left(p_t^j q_{i,t}^j - \sum_m w_{m,t} x_{i,m,t}^j\right) \\
\text{s.t. } \textstyle\sum_j d_{i,t}^j &= 1 \\
q_{i,t}^j &= Q_j\left(\mathbf{x}_{i,t}^j\right) \\
x_{i,m,t}^j &\geq 0 \\
j &= 0, \ldots, J
\end{aligned} \tag{2.1}$$

where $d_{i,t}^j = 1$ if crop j is planted in year t and 0 otherwise; p_t^j is the expected price for crop j in year t; $q_{i,t}^j$ is the expected yield for crop j on field i in year t; $x_{i,m,t}^j$ is the level of input m used in the production of crop j in year t on field i with associated cost $w_{m,t}$; and Q_j is a concave production function for crop j.

Growing crops in rotation can have substantial benefits. For example, Karlen, Berry, Colvin, and Kanwar (1991) found crop yields and nitrogen use efficiencies improved under a corn-soybean rotation as opposed to a continuous cropping system. Using data from experiment stations, Rozeboom et al. (2009) showed that the major crops in Kansas — corn, sorghum, soybeans, and wheat — can all see yield benefits through a rotational system. With wheat, for example, 10-year average yields were higher in a no-till system following either corn or soybeans than following wheat by roughly 8 to 10/bu./ac. (Rozeboom et al., 2009). Three-year average sorghum yields, meanwhile, were approximately 20/bu./ac. higher following wheat than following sorghum (Rozeboom et al., 2009). Thus, one component of the rotational effect is assumed to be a yield boost, β, which is defined as an additive component to crop yields, following Hendricks, Smith, and Sumner (2014); Hennessy (2006). Yield boosts are assumed to have a one-year memory and be dependent on both the past and current crop, e.g., the yield boost for soybeans in year t following corn is not the same as the yield boost following wheat. Thus, we represent the yield boost received by crop j on field i in year t as $\beta_{i,t}^{\mathbf{r}}$ where $\mathbf{r}$ denotes the two-year crop sequence $\langle y_{t-1}, y_t\rangle$, and $y_{t-\ell} \in \{j = 0, \ldots, J\}$ for $\ell \in \{0, 1\}$. The yield boost received by crop j on field i in year t can then be represented by $\beta_{i,t}^{\mathbf{r}_j}$, where $\mathbf{r}_j \in R_j$ is the crop sequence $\langle y_{t-1}, j\rangle$ and R_j denotes the set of all $J+1$ possible crop sequences for which $y_t = j$. Using this, the second constraint in problem 2.1 can be written as:

$$q_{i,t}^{j} = Q_j\left(\mathbf{x}_{i,t}^{j}\right) + \beta_{i,t}^{\mathbf{r}_j}. \tag{2.2}$$

and substituting equation 2.2 into the objective function, we can rewrite problem 2.1 as

$$\begin{aligned} \max \Pi_{i,t} &= \textstyle\sum_j d_{i,t}^{j}\left[p_t^{j}\left(Q_j\left(\mathbf{x}_{i,t}^{j}\right) + \beta_{i,t}^{\mathbf{r}_j}\right) - \sum_m w_{m,t} x_{i,m,t}^{j}\right] \\ \text{s.t. } \textstyle\sum_j d_{i,t}^{j} &= 1 \\ x_{i,m,t}^{j} &\geq 0 \\ j &= 0, \ldots, J \end{aligned} \quad . \tag{2.3}$$

Besides past cropping history, current yields are assumed to be a function of other uncontrollable and locational variables such as soil characteristics $(\mathbf{s}_{i,t})$ and climate variables $(\mathbf{c}_{i,t})$, as well as the level of inputs used $\left(\mathbf{x}_{i,t}^{j}\right)$. Incorporating these elements, the yield production function for crop j in year t becomes:

$$q_{i,t}^{j} = Q_j\left(\mathbf{x}_{i,t}^{j}, \mathbf{s}_{i,t}, \mathbf{c}_{i,t}\right) + \beta_{i,t}^{\mathbf{r}_j}. \tag{2.4}$$

Soils variables could be those related to overall soil health and productivity, such as organic matter content or drainage, and may also depend on past cropping and management decisions. Weather variables are assumed to be a key factor in determining yields and could include precipitation measures (e.g., Schlenker and Roberts (2009); Hendricks et al. (2014); Isik and Devadoss (2006)), growing degree days (e.g., C. Reynolds et al. (2000)), etc. Many of the elements in $\mathbf{c}_{i,t}$, such as growing season precipitation, will be unknown to farmers at the time decisions are made. This creates uncertainty about yield outcomes and profits. Thus, it is assumed that farmers form expectations on what will occur and factor those expectations into their land-use decisions. Additionally, it is assumed that farmers have adaptive expectations, i.e., their expectations may change after the current year's weather has been realized and taken into consideration.

Incorporating equation 2.4, the farmer's problem is now:

$$\begin{aligned} \max \Pi_{i,t} &= \textstyle\sum_j d_{i,t}^j \left(p_t^j \left[Q_j \left(\mathbf{x}_{i,t}^j, \mathbf{s}_{i,t}, \mathbf{c}_{i,t} \right) + \beta_{i,t}^{\mathbf{r}j} \right] - \sum_m w_{m,t} x_{i,m,t}^j \right) \\ \text{s.t. } \textstyle\sum_j d_{i,t}^j &= 1 \\ x_{i,m,t}^j &\geq 0 \\ j &= 0, \ldots, J \end{aligned} \quad . \tag{2.5}$$

Problem 2.5 can be solved sequentially by first solving for the maximum profit under each crop and then setting $d_{i,t}^j = 1$, where j denotes the crop with the highest expected profit. Ignoring the terms in $\mathbf{d}_{i,t} = \left[d_{i,t}^0, d_{i,t}^1, \ldots, d_{i,t}^J \right]$, the only decision variables for the farmer are the input levels to be used under each crop. With this in mind, first-order conditions for the sub-problem associated with crop j are given by:

$$p_t^j \frac{\partial Q_j}{\partial x_{i,m,t}^j} = w_{m,t} \tag{2.6}$$

or

$$\frac{\partial Q_j}{\partial x_{i,m,t}^j} = \frac{w_{m,t}}{p_t^j}. \tag{2.7}$$

A system of equations is obtained from the first order conditions for each crop $j = 0, \ldots, J$. Following Hennessy (2006), we denote the solutions to these systems as:

$$x_{i,m,t}^{j*} = g_m^j \left(\mathbf{w}_{i,t}^j, \mathbf{s}_{i,t}, \mathbf{c}_{i,t} \right), \tag{2.8}$$

for $j = 0, \ldots, J$, where $\mathbf{w}_{i,t}^j$ is a vector of price ratios with individual element $w_{i,m,t}^j = \frac{w_{i,m,t}}{p_t^j}$.

Using equation 2.8, maximum profit for crop j on field i in year t can be expressed as:

$$\Pi_{i,t}^j = p_t^j \left[Q_j \left(\mathbf{g}^j \left(\mathbf{w}_{i,t}^j, \mathbf{s}_{i,t}, \mathbf{c}_{i,t} \right), \mathbf{s}_{i,t}, \mathbf{c}_{i,t} \right) + \beta_{i,t}^{\mathbf{r}j} \right] - \sum_m w_{m,t} g_m^j \left(\mathbf{w}_{i,t}^j, \mathbf{s}_{i,t}, \mathbf{c}_{i,t} \right) \tag{2.9}$$

The farmer then decides to plant crop j if the following condition holds:

$$\Pi_{i,t}^{j} - \Pi_{i,t}^{k} > 0 \text{ for all } k \neq j \tag{2.10}$$

or

$$\left[p_t^j\left(Q_j + \beta_{i,t}^{\mathbf{r}_j}\right) - \sum_m w_{m,t} g_m^j\right] - \left[p_t^k\left(Q_k + \beta_{i,t}^{\mathbf{r}_k}\right) - \sum_m w_{m,t} g_m^k\right] > 0 \ \forall k \neq j, \tag{2.11}$$

where $Q_j = Q_j\left(\mathbf{g}^j\left(\mathbf{w}_{i,t}^j, \mathbf{s}_{i,t}, \mathbf{c}_{i,t}\right), \mathbf{s}_{i,t}, \mathbf{c}_{i,t}\right)$ and $g_m^j = g_m^j\left(\mathbf{w}_{i,t}^j, \mathbf{s}_{i,t}, \mathbf{c}_{i,t}\right)$. Collecting like terms and suppressing the condition $\forall k \neq j$, condition 2.11 can be rewritten as:

$$p_t^j\left(Q_j + \beta_{i,t}^{\mathbf{r}_j}\right) - p_t^k\left(Q_k + \beta_{i,t}^{\mathbf{r}_k}\right) + \sum_m w_{m,t}\left(g_m^k - g_m^j\right) > 0. \tag{2.12}$$

Now assume that the yield boost term can be decomposed as follows:

$$\beta_{i,t}^{\mathbf{r}_j} = \alpha_{0,i,t}^j + \sum_{s=0}^{J} \alpha_{1,i,t}^s d_{i,t-1}^s, \tag{2.13}$$

to capture dependence with the past crop sequence. After substitution of equation 2.13, the condition in 2.12 becomes:

$$\begin{aligned} p_t^j\left(Q_j + \alpha_{0,i,t}^j + \textstyle\sum_{s=0}^J \alpha_{1,i,t}^s d_{i,t-1}^s\right) - p_t^k\left(Q_k + \alpha_{0,i,t}^k + \textstyle\sum_{s=0}^J \alpha_{1,i,t}^s d_{i,t-1}^s\right) + \\ \textstyle\sum_m w_{m,t}\left(g_m^k - g_m^j\right) > 0. \end{aligned} \tag{2.14}$$

Because it is assumed that the $\alpha_{1,i,t}^s$ terms are independent of the crop choice in period t, the condition in 2.14 simplifies to:

$$\begin{aligned} p_t^j\left(Q_j + \alpha_{0,i,t}^j\right) - p_t^k\left(Q_k + \alpha_{0,i,t}^k\right) + \left(p_t^j - p_t^k\right) \textstyle\sum_{s=0}^J \alpha_{1,i,t}^s d_{i,t-1}^s + \\ \textstyle\sum_m w_{m,t}\left(g_m^k - g_m^j\right) > 0. \end{aligned} \tag{2.15}$$

2.4 Dynamic-Multinomial Logit with Random Effects

The realized profits, $\Pi_{i,t}^{j}$, contain several components which are unknown to or unobserved by the researcher (and possibly the farmer), such as the yield function or yield boosts. Additionally, at the time the decision is made, there may be uncertainty regarding some variables (e.g. prices or weather). Thus, $\Pi_{i,t}^{j},\ j=0,\ldots,J$ are random variables. Assuming $\Pi_{i,t}^{j}$ can be decomposed into a known component, $\pi_{i,t}^{j}$, and a random component, $\varepsilon_{i,t}^{j}$, condition 2.10 becomes

$$\left(\pi_{i,t}^{j}+\varepsilon_{i,t}^{j}\right)-\left(\pi_{i,t}^{k}+\varepsilon_{i,t}^{k}\right)>0 \text{ for all } k\neq j \tag{2.16}$$

or

$$\pi_{i,t}^{j}-\pi_{i,t}^{k}>\varepsilon_{i,t}^{k}-\varepsilon_{i,t}^{j} \text{ for all } k\neq j. \tag{2.17}$$

Based on condition 2.17, condition 2.15 becomes

$$\begin{aligned}&p_t^j\left(Q_j+\alpha_{0,i,t}^{j}\right)-p_t^k\left(Q_k+\alpha_{0,i,t}^{k}\right)+\\&\left(p_t^j-p_t^k\right)\textstyle\sum_{s=0}^{J}\alpha_{1,i,t}^{s}d_{i,t-1}^{s}+\sum_m w_{m,t}\left(g_m^k-g_m^j\right)>\varepsilon_{i,t}^{k}-\varepsilon_{i,t}^{j}.\end{aligned} \tag{2.18}$$

It is assumed that the error terms, $\varepsilon_{i,t}^{j}$, are independent and identically distributed with mean zero.

Based on condition 2.18, the probability that field i is planted to crop j in year t, i.e., $y_{i,t}=j \Rightarrow d_{i,t}^{j}=1$ and $d_{i,t}^{k}=0$ for $k\neq j$, can be derived as:

$$\begin{aligned}P\left(y_{i,t}=j\right)=\ &P\Big[p_t^j\left(Q_j+\alpha_{0,i,t}^{j}\right)-p_t^k\left(Q_k+\alpha_{0,i,t}^{k}\right)+\left(p_t^j-p_t^k\right)\textstyle\sum_{s=0}^{J}\alpha_{1,i,t}^{s}d_{i,t-1}^{s}+\\&\textstyle\sum_m w_{m,t}\left(g_m^k-g_m^j\right)>\varepsilon_{i,t}^{k}-\varepsilon_{i,t}^{j}\ \forall k\neq j\Big],\end{aligned}\ . \tag{2.19}$$

Assuming a linear-in-parameters functional form for the left-hand side of the inequality in condition 2.19, the condition can be represented by:

$$P\left(y_{i,t}=j\right)=P\left(\mathbf{z}_{i,t}'\boldsymbol{\beta}_j>\zeta_{k,j,t}\right)\ \forall k\neq j, \tag{2.20}$$

where

$$\mathbf{z}_{i,t} = [\mathbf{p}_{i,t}\ \mathbf{w}_{i,t}\ \mathbf{s}_{i,t}\ \mathbf{c}_{i,t}\ \mathbf{d}_{i,t-1}]', \tag{2.21}$$

$$\zeta_{j,k,t} = \varepsilon_{k,t} - \varepsilon_{j,t}, \tag{2.22}$$

and $\boldsymbol{\beta}_j$ $(j = 0, \ldots, J)$ is a set of parameters — specific to land-use j — to be estimated. An index or predictor function that is also linear in the explanatory variables can be viewed as a first order Taylor series approximation of the difference in expected profits for two land uses.

Assuming also that ε_j and ε_k in equation 2.22 are distributed *IID* Gumbel, the probabilities can be estimated with a standard multinomial logit (SML) model. For the SML, the probability that field i is allocated to use j in year t is given by

$$P^j_{i,t} = \frac{\exp\left(\mathbf{z}'_{i,t}\boldsymbol{\beta}_j\right)}{\sum_{h=0}^{J} \exp\left(\mathbf{z}'_{i,t}\boldsymbol{\beta}_h\right)}. \tag{2.23}$$

In order to secure identifiability of parameters, the parameters for land use $j = 0$ are normalized to zero (i.e., $\boldsymbol{\beta}_0 = \mathbf{0}$), and the probabilities for $j = 1, \ldots, J$ become:

$$P^j_{i,t} = \frac{\exp\left(\mathbf{z}'_{i,t}\boldsymbol{\beta}_j\right)}{1 + \sum_{h=1}^{J} \exp\left(\mathbf{z}'_{i,t}\boldsymbol{\beta}_h\right)}. \tag{2.24}$$

A number of potential issues arise with the basic SML model presented in equation 2.23. First, this model imposes the "independence from irrelevant alternatives" (IIA) property. When IIA holds, it implies that, for a particular farmer, the odds-ratio between to competing land uses is unaffected by the presence of other alternatives. This can be seen in equation 2.25 below:

$$\frac{P^j_{i,t}}{P^h_{i,t}} = \frac{\exp\left(\mathbf{z}'_{i,t}\boldsymbol{\beta}_j\right)}{\exp\left(\mathbf{z}'_{i,t}\boldsymbol{\beta}_h\right)}. \tag{2.25}$$

The IIA property, while convenient for estimation, is not always plausible. A number of alternatives exist that can relax this assumption, such as the nested logit, multinomial probit[1],

[1]In practice, the multinomial probit is rarely used due to estimation obstacles; see W. H. Greene (2012, p. 811) for an overview.

and mixed logit models[2].

A second concern is unobserved heterogeneity across fields. With the current data set, this risk is compounded due to the absence of farmer-specific data, and thus the inability to control for individual farmer specific characteristics such as management ability, resource constraints, etc. In the presence of unobserved heterogeneity, equation 2.23 can be written as:

$$P_{i,t}^{j} = \frac{\exp\left(\mathbf{z}_{i,t}'\boldsymbol{\beta}_j + a_{i,j}\right)}{1+\sum_{h=1}^{J}\exp\left(\mathbf{z}_{i,t}'\boldsymbol{\beta}_h + a_{i,h}\right)} \tag{2.26}$$

where $a_{i,j}$, $j = 1, \ldots, J$ are the unobserved effects. When panel data is available, the problem of unobserved heterogeneity is often treated with a random- or fixed-effects approach. With the fixed-effects approach, each unobserved effect is treated as a parameter to be estimated, which results in a potential incidental parameters problem. With the random-effects approach, it is assumed that: $a_{i,j} \mid \mathbf{z}_i \sim N\left(0, \sigma_j^2\right)$,

$$P_{i,j}\left(y_{i,t} = j \mid \mathbf{z}_i, a_{i,j}\right) \equiv P_{i,j}\left(y_{i,t} = j \mid \mathbf{z}_{i,1}, \ldots, \mathbf{z}_{i,T}, a_{i,j}\right) = P_{i,j}\left(y_{i,t} = j \mid \mathbf{z}_{i,t}, a_{i,j}\right), \tag{2.27}$$

and

$$y_{i,1}, \ldots, y_{i,T} \text{ are independent conditional on } \mathbf{z}_i, \mathbf{a}_i. \tag{2.28}$$

Assumption 2.27 implies that, conditional on $(\mathbf{z}_{i,t}, a_i)$, $d_{i,t} \perp d_{i,s}$ for $t \neq s$; or in other words, the assumption rules out models with lagged-dependent variables (Wooldridge, 2010). Following W. Greene (2012), it is typically assumed that

$$\mathbf{a}_i = \boldsymbol{\Gamma}\mathbf{v}_i \tag{2.29}$$

where $\boldsymbol{\Gamma}$ is a $J \times J$ lower-triangular matrix and v_i is a $J \times 1$ vector distributed $N\left(\mathbf{0}, \mathbf{I}_j\right)$. Under this specification, $\mathbf{a}_i \sim N\left(\mathbf{0}, \boldsymbol{\Gamma}\boldsymbol{\Gamma}'\right)$. This assumes the individual (field) effects are correlated across equations. To assume no correlation across the random effects, the off diagonal elements can be set to zero.

[2]Also referred to as random parameters logit.

For a panel data model where unobserved heterogeneity is assumed, the inclusion of the lagged-dependent variable, $\mathbf{d}_{i,t-1}$, on the right-hand side of equation 2.26 creates an additional concern. In this case, with a slight notational change, condition 2.20 may be written as:

$$P\left(y_{i,t}=j\right)=P\left(\mathbf{z}_{i,t}'\boldsymbol{\beta}_j+\mathbf{d}_{i,t-1}'\boldsymbol{\rho}+u_{i,j}>\zeta_{k,j,t}\right)\ \forall k\neq j \tag{2.30}$$

where $u_{i,j}$ now represents the unobserved effect and $\boldsymbol{\rho}_j$ are the parameters associated with the lagged-dependent variables. Similarly, the probability in equation 2.26 may be written as:

$$P_{i,t}^{j}=\frac{\exp\left(\mathbf{z}_{i,t}'\boldsymbol{\beta}_j+\mathbf{d}_{i,t-1}'\boldsymbol{\rho}_j+u_{i,j}\right)}{1+\sum_{h=1}^{J}\exp\left(\mathbf{z}_{i,t}'\boldsymbol{\beta}_h+\mathbf{d}_{i,t-1}'\boldsymbol{\rho}_h+u_{i,h}\right)} \tag{2.31}$$

or more compactly as:

$$P_{i,t}^{j}=\Lambda\left(\mathbf{z}_{i,t},\mathbf{d}_{i,t-1},\mathbf{u}_i;\boldsymbol{\beta},\boldsymbol{\rho}\right), \tag{2.32}$$

where Λ denotes the multinomial logistic cumulative density function. Then, suppressing the field subscript i, the joint density for a field is given by:

$$f\left(y_1,\ldots,y_T\mid y_{t-1},\ldots,y_0,\mathbf{z},\mathbf{u};\boldsymbol{\beta},\boldsymbol{\rho}\right)=\prod_{t=1}^{T}\prod_{j=0}^{J}\left[\Lambda\left(\mathbf{z}_{i,t},\mathbf{d}_{i,t-1},\mathbf{u}_i;\boldsymbol{\beta},\boldsymbol{\rho}\right)\right]^{d_t^j}. \tag{2.33}$$

Wooldridge (2010) notes that, for the binary case ($j\in(0,1)$), the presence of the unobserved effect does not allow for a log-likelihood function that can be used to consistently estimate $\boldsymbol{\beta}$ and $\boldsymbol{\rho}$. Instead, it is suggested that the unobserved effects be integrated out of the distribution, which creates a new issue of how to deal with the initial observations, y_0 (Wooldridge, 2010). This is generally referred to as the initial conditions problem. W. H. Greene (2012) notes that the initial conditions can have a crucial impact on the entire path of outcomes and, additionally, standard estimators are no longer consistent. To address the initial conditions problem, an alternative model was proposed by Wooldridge (2005) for the binary case. Though other techniques are available, such as that from Heckman (1981), the approach from Wooldridge (2005) is appealing as it results in a specification which is estimable by standard econometric software.

With respect to the need to integrate out the unobserved effects, Wooldridge (2005) noted a popular approach was to specify a density $g\left(y_0 \mid \mathbf{z}, \mathbf{u}\right)$ to yield:

$$f\left(y_0, y_1, \ldots, y_T \mid \mathbf{z}, \mathbf{u}; \boldsymbol{\beta}, \boldsymbol{\rho}\right) = \prod_{t=1}^{T} \prod_{j=0}^{J} \left[\Lambda\left(\mathbf{z}_{i,t}, \mathbf{d}_{i,t-1}, \mathbf{u}_i; \boldsymbol{\beta}, \boldsymbol{\rho}\right)\right]^{d_t^j} g\left(y_0 \mid \mathbf{z}, \mathbf{u}\right). \tag{2.34}$$

Equation 2.34 can then be integrated with respect to a density $h\left(\mathbf{u} \mid \mathbf{z}\right)$ — specified by the researcher — to obtain $f\left(y_0, y_1, \ldots, y_T \mid \mathbf{z}\right)$ (Wooldridge, 2005). While the resulting density can then be estimated via maximum likelihood, the density $g\left(y_0 \mid \mathbf{z}, \mathbf{u}\right)$ is extremely difficult — if not impossible — to define (Wooldridge, 2010). Alternatively, Wooldridge (2005) suggests using $f\left(y_1, y_2, \ldots, y_T \mid \mathbf{z}, y_0\right)$ because $f\left(y_1, y_2, \ldots, y_T \mid y_0, \mathbf{z}, \mathbf{u}\right)$ has already been specified using equation 2.33. Then, all that must be done is to specify a density for $\mathbf{u}$ conditional on y_0 and $\mathbf{z}$.

Letting $P\left(\mathbf{u} \mid y_0, \mathbf{z}; \boldsymbol{\alpha}\right) = h\left(\mathbf{u} \mid y_0, \mathbf{z}; \boldsymbol{\alpha}\right)$, the density of $\mathbf{y} \mid y_0, \mathbf{z}$ is given by (Wooldridge, 2005):

$$\int_{\mathbb{R}^J} f\left(y_1, y_2, \ldots, y_T \mid \mathbf{z}, y_{t-1}, \ldots, y_1, y_0, \mathbf{u}; \boldsymbol{\beta}\right) h\left(\mathbf{u} \mid y_0, \mathbf{z}; \boldsymbol{\alpha}\right) \eta\left(d\mathbf{u}\right), \tag{2.35}$$

which through substitution of equation 2.33 becomes:

$$\int_{\mathbb{R}^J} \prod_{t=1}^{T} \prod_{j=0}^{J} \left[\Lambda\left(\mathbf{z}_{i,t}, \mathbf{d}_{i,t-1}, \mathbf{u}_i; \boldsymbol{\beta}, \boldsymbol{\rho}\right)\right]^{d_t^j} h\left(\mathbf{u} \mid y_0, \mathbf{z}; \boldsymbol{\alpha}\right) \eta\left(d\mathbf{u}\right). \tag{2.36}$$

It is then possible to specify $h\left(\cdot\right)$ in such a way that the model can be estimated using a standard random-effects approach. Common specifications for $h\left(\cdot\right)$ include:

$$u_{i,j} \mid y_{i,0}, \mathbf{z}_i \sim Normal\left(\alpha_{0,j} + \mathbf{d}_{i,0}'\boldsymbol{\alpha}_{1,j} + \mathbf{z}_i'\boldsymbol{\alpha}_{2,j}, \sigma_j^2\right) \tag{2.37}$$

or

$$u_{i,j} \mid y_{i,0}, \mathbf{z}_i \sim Normal\left(\alpha_{0,j} + \mathbf{d}_{i,0}'\boldsymbol{\alpha}_{1,j} + \bar{\mathbf{z}}_i'\boldsymbol{\alpha}_{2,j}, \sigma_j^2\right) \tag{2.38}$$

where $\mathbf{z}_i = \left[\mathbf{z}_{i,0}' \; \mathbf{z}_{i,1}' \; \cdots \; \mathbf{z}_{i,T}'\right]'$ is the full history of the explanatory variables and $\bar{\mathbf{z}}_i$ are the means of the explanatory variables for cross-sectional unit i and $a_{i,j}$ is specified as in equation 2.29. Under these specifications, the random effects can then be written as

$$u_{i,j} = \alpha_{0,j} + \mathbf{d}'_{i,0}\boldsymbol{\alpha}_{1,j} + \mathbf{z}'_i\boldsymbol{\alpha}_{2,j} + a_{i,j} \tag{2.39}$$

or

$$u_{i,j} = \alpha_{0,j} + \mathbf{d}'_{i,0}\boldsymbol{\alpha}_{1,j} + \bar{\mathbf{z}}'_i\boldsymbol{\alpha}_{2,j} + a_{i,j} \tag{2.40}$$

Equation 2.34 along with either assumption 2.37 or 2.38 is referred to as a dynamic-multinomial logit with random effects model (DML-RE).

Farmer specific weather expectations and the process by which they form these expectations are likely unique to the individual. In other words, the weather-expectation process is one justification for the unobserved heterogeneity assumption. If this is individual specific, then two farmers may hold different weather expectations for the upcoming season even if they have experienced identical weather histories. Under this scenario, the two farmers may make different land-use decisions in any given year, even if all other variables are identical. This then, would appear to justify the approach from Wooldridge (2005) to specify the unobserved-effects distribution as a function of the initial conditions.

2.5 Data

Analysis for this study is based on a subset of Kansas counties, which are depicted in figure 2.1. The primary reason for choosing these counties is that their combined boundaries — more or less — encompass two central Kansas watersheds which are the subject of ongoing research. From this starting point, the original counties were adjusted slightly. Specifically, Sheridan County was dropped and Ottawa County was brought in. This was done to limit the effects of irrigation, which are not captured in the empirical model. For the included counties, irrigated cropland averaged about 2% of total cropland from 2000 to 2015; Sheridan County averaged about 21% (United States Geological Survey, 2017). An additional motivation for this set of counties comes from the fact that Kansas exhibits a strong precipitation gradient moving from east to west. This gradient, depicted in figure 2.2, represents a difference of about 11 inches per year in the 30-year average over the period 1981-2010.

2.5.1 Land-use Decision Variables

Land-use observations are based on a detailed, field-level database for the state of Kansas. Field boundaries are provided by a large spatial-polygons file and land-use decisions within each boundary are based upon the United States Department of Agriculture, National Agricultural Statistics Service (2016) Cropland Data Layer (CDL) and Farm Service Agency (FSA) data for the years 2003 to 2012. The study region depicted in figure 2.1 consists of 157,212 unique fields, developed by the Kansas Biological Survey. When observations in the study region are limited to the land uses of interest — corn, sorghum, soybeans, wheat, and an "other" category— the number of fields drops to 45,542. The other category denotes a land use of double cropping (5.5%), alfalfa (27.1%), or fallow (67.4%). The locations of the included fields are indicated in figure 2.3. Excluded fields include those that were devoted to grassland or minor crops and those that were developed land or water. One year is set aside as the initial period, leaving a balanced-panel set of 409,878 observations for estimation.

Dependent variables are the five land uses described above and are denoted COR_t, SOR_t, SOY_t, WHT_t, and OTH_t. Summary statistics for these categories can be found in tables 2.1 and 2.2. Wheat was the dominant land use over the period examined, with an average of 54% of fields and 51% of acres devoted to wheat each year. With an average share of 5% of fields per year, corn was generally the most infrequent land use. In terms of acreage, the lowest share generally went to soybeans, which had an average acre share of 5%. In terms of both number of fields and acres, corn shares generally trended up while wheat shares trended down. Descriptions and averages for the lagged land-use — COR_{t-1}, SOR_{t-1}, SOY_{t-1}, WHT_{t-1}, and OTH_{t-1} — and the initial conditions variables — COR_0, SOR_0, SOY_0, WHT_0, and OTH_0 — are presented in table 2.3. Wheat again dominated, with 62% of the fields devoted to wheat in period $t = 0$ (i.e., year 2003) and corn was again the least represented with only 2% of fields in $t = 0$.

2.5.2 Expected Output Prices

The price received at harvest (or whenever a harvested crop is sold) is unknown at the time the land-use decision is made. For this reason, expected prices for a given crop are used. These variables are meant to represent the price a farmer expects to receive at harvest given the information available at the time a crop is planted. Expected prices were created following the approach of Hendricks et al. (2014). Under this approach, expected prices are set equal to a futures price at the time of planting plus an expectation of the basis at harvest:

$$E\left(p_t^j\right) = FP_t^j + E\left(B_t^j\right). \tag{2.41}$$

The futures price component $\left(FP_t^j\right)$ was calculated from daily futures price data for each crop. The daily futures prices were averaged across the "planting months" for a given crop: March and April for corn, sorghum, and soybeans and September and October for wheat. The futures price for sorghum was set equal to that of corn because sorghum futures prices were not available for this time period. The expected basis was set equal to the harvest-time basis from the previous harvest. For corn, sorghum, and soybeans, this comes from the previous calendar year, for wheat it comes from the same calendar year. Spot price data for basis calculations came from 961 elevator locations across Kansas.

Expected prices were calculated for each of the 961 elevator locations. Because some elevators were assigned the same latitude and longitude coordinates, e.g., all elevators in Salina, Kansas, there may be more than one "nearest" elevator for a given field. Thus, the expected-crop price for a given field was set equal to the mean price of the nearest elevators for which data was available. Data were not available for every crop-year-elevator combination, and so the expected price for a particular field across crops may come from different elevator locations both within and across years. Summary statistics for the included expected-price variables — $PCORN$, $PSOY$, and $PWHEAT$ — can be found in table 2.4.

It is recognized that input prices are important in determining the net returns to the different land uses. However, their inclusion in the empirical model resulted in a singular covariance matrix and they were thus excluded from the final analysis. However, Hendricks et al. (2014) show that changes in fertilizer prices have a much smaller impact on relative crop returns than output prices and that their results were robust to the exclusion of input prices from analysis. Additionally, Hendricks et al. (2014) noted that with a short panel — 11 years in their case — the inclusion of fertilizer creates limitations in the ability to identify the impact of crop prices.

2.5.3 Soils Variables

Soils variables were calculated using the USDA Natural Resources Conservation Service Gridded Soil Survey Geographic (gSSURGO) data for Kansas (United States Department of Agriculture, National Resources Conservation Service, 2016). The gSSURGO geodatabase (available on-line at https://datagateway.nrcs.usda.gov) contains shapefiles that delineate soil-type boundaries. Soil types are identified by a map-unit key (MUKEY), of which 5,795 are represented in Kansas. Separate files within the gSSURGO geodatabase allow a MUKEY to be linked to physical characteristics for that soil type. Two soil-property variables were included: *CLAY* and *SILT*, which give the shares of clay and silt found in a field's soil. The remaining share is composed of sand, which is excluded from the models. Often, more than one soil type was present within a field boundary, so the variables here represent a weighted average across soil types. Summary statistics for *CLAY* and *SILT* are presented in table 2.4.

2.5.4 Weather Variables

Included weather variables were calculated using daily data provided by Schlenker and Roberts (2009) (available at http://www.wolfram-schlenker.com/dailyData.html). The data sets provided by Schlenker and Roberts (2009) were created using data from the PRISM Weather Group. The daily weather data provided by PRISM are interpolated values based on weather stations located throughout the United States and are provided for gridded units that are roughly 16 km^2 in size. Individual fields were assigned weather data based on which

PRISM grid cell they primarily lie within. Summary statistics are provided in table 2.4.

The first weather variable — *PRECSUM* — measures precipitation during the current year for the months of April, May, and June. This measure is the average daily precipitation in millimeters (mm). The average of 2.63 mm per day per week corresponds to roughly 9.4 total inches over these three months. *PRECSUM* is intended to primarily capture prevented planting effects and thus is expected to have an impact on only the spring planted crops: corn, sorghum, and soybeans.

The remaining weather variables can be viewed as falling into one of two categories: measures of average-historical precipitation or of average-historical-precipitation variability. Each group includes two variables, denoted $M3$ and $M10$ for the first and $S3$ and $S10$ for the second. $M3$ and $M10$ provide three- and ten-year averages, respectively, of average weekly precipitation, defined in terms of an average daily precipitation. The $S3$ and $S10$ variables provide the standard deviation of the weekly precipitation, averaged over either the past three or ten years. For example, if the average daily precipitation per week for the past three years are given by $PR_{t-1} = 2$ mm, $PR_{t-2} = 2.5$ mm, and $PR_{t-3} = 3$ mm, then $M3_t = \frac{1}{3}(2 + 2.5 + 3) = 2.5$ mm. This is simply a rescaling of total precipitation over the six months. If the standard deviation of the average daily precipitation across weeks are given by $S_{t-1} = 2.5$, $S_{t-2} = 3$, and $S_{t-3} = 3$, then $S3_t = \frac{1}{3}(2.5 + 3 + 3.5) = 3$. Both sets of variables are based on a loosely-defined growing season consisting of April through September. The values for $M3$ and $M10$ in table 2.4 correspond (approximately) to three- and ten-year moving averages of 0.10 inches of precipitation per day per week. Over the course of these six months, these values equate to a little more than 18 inches of total precipitation. The minimum and maximum values indicate a range of between approximately 11.8 and 26.9 inches for $M3$ and 13.0 and 23.3 inches for $M10$.

To provide context as to how these levels may impact land-use decisions, a table from Stone and Schlegel (2006) is reproduced in table 2.5. The table from Stone and Schlegel (2006) is based on research conducted in western Kansas and is thus applicable to the current study region. The table indicates lower threshold evapotranspiration (ET) levels for soybeans and grain sorghum of 7.8 inches and 6.9 inches respectively, where threshold ET is defined as the ET level below which seed yield is zero (Stone & Schlegel, 2006). Grain sorghum is shown to have the lowest full-season requirements and also one of the more responsive yield functions with respect to inches of water. Corn is seen to require the most water in terms of full season needs and threshold ET. The estimates from Stone and Schlegel (2006) offer validation to the conclusions reached by Staggenborg et al. (2008).

Ideally, additional weather controls would have been included in the analysis, such as temperature measures and interactions between precipitation and temperature or between precipitation variables, e.g., $M10 \times S10$. Due to estimation difficulties, the four described above were maintained. It could also be argued to use different months to define a growing season. The variables described were chosen as it is believed they represent important months for each of the five land uses while not being of too long or too short a duration.

2.6 Post-Estimation Inferences

2.6.1 Average Partial Effects

Of particular interest in this analysis are the average partial effects (APEs) associated with the historical precipitation variables, $M3$, $M10$, $S3$, and $S10$. In general, for a continuous variable, the partial effect for an individual on the probability of choosing land-use j is given by

$$\frac{\partial P^j}{\partial x_k} = P^j \left(\beta_{j,k} - \sum_{h=1}^{J} \beta_{h,k} P^h \right), \tag{2.42}$$

where the individual and time subscripts have been dropped. For the weather-variable partial effects as given in equation 2.42 — the typical output from econometric software — a special interpretation is needed. For an increase in $M3$, the typical partial effect implies that there is no change in $M10$, and so there is an implicit decrease in the precipitation in years four through ten. Thus, the standard partial effect for $M3$ can be thought of as a reallocation of precipitation from years four through ten to years one through three in such a way that $M10$ remains unchanged. Under this interpretation, the partial effect of $M3$ can be viewed as the impact on the probability of observing a particular land use from increasing the relative "wetness" of the past three years compared to years four through ten. Similarly, the partial effect on $M10$ can be viewed as an increase in the relative "dryness" of years one through three. In this case, however, the interpretation can be arrived at without assuming any change in precipitation for the previous three years by assuming the change in $M10$ accrues only in years four through ten.

An alternative approach is to look at the partial effects associated with a change in just one particular year. Let β_{M3} and β_{M10} denote the parameters associated with $M3$ and $M10$. Then, since $M3_t = \frac{1}{3}\sum_{\ell=1}^{3} PR_{t-\ell}$ and $M10_t = \frac{1}{10}\sum_{\ell=1}^{10} PR_{t-\ell}$, the partial effect associated with $PR_{t-\ell}$ is given by:

$$\frac{\partial P^j}{\partial PR_{t-\ell}} = P^j \left[\left(\frac{\beta_{j,M10}}{10} \right) - \sum_{h=1}^{J} \left(\frac{\beta_{h,M10}}{10} \right) P^h \right] \tag{2.43}$$

or

$$\frac{\partial P^j}{\partial PR_{t-\ell}} = \frac{1}{10} P^j \left[\beta_{j,M10} - \sum_{h=1}^{J} \beta_{h,M10} P^h \right] = \frac{1}{10} \frac{\partial P^j}{\partial M10} \tag{2.44}$$

for $\ell = 4, \ldots, 10$, and

$$\frac{\partial P^j}{\partial PR_{t-\ell}} = P^j \left[\left(\frac{\beta_{j,M3}}{3} + \frac{\beta_{j,M10}}{10} \right) - \sum_{h=1}^{J} \left(\frac{\beta_{h,M3}}{3} + \frac{\beta_{h,M10}}{10} \right) P^h \right], \tag{2.45}$$

which is equivalent to:

$$\frac{\partial P^j}{\partial PR_{t-\ell}} = \frac{1}{3} \frac{\partial P^j}{\partial M3} + \frac{1}{10} \frac{\partial P^j}{\partial M10} \tag{2.46}$$

for $\ell = 1, 2, 3$, where $\frac{\partial P^j}{\partial M3}$ and $\frac{\partial P^j}{\partial M10}$ are the standard partial effects associated with $M3$ and $M10$. Then, for an arbitrary change $\Delta PR_{t-\ell} = \gamma$, the impact on P^j can be approximated by:

$$\frac{\partial P^j}{\partial PR_{t-\ell}} = \frac{\gamma}{10} \frac{\partial P^j}{\partial M10} \tag{2.47}$$

for $\ell = 4, \ldots, 10$ and

$$\frac{\partial P^j}{\partial PR_{t-\ell}} = \frac{\gamma}{3} \frac{\partial P^j}{\partial M3} + \frac{\gamma}{10} \frac{\partial P^j}{\partial M10} \tag{2.48}$$

for $\ell = 1, 2, 3$. To get an idea as to how important the past three years of weather are compared to weather that occurred four to ten years ago, we can use the following equations:

$$\frac{\partial P^j}{\partial PR_{t-\ell}} = \frac{7\gamma}{10}\frac{\partial P^j}{\partial M10} \tag{2.49}$$

for $\ell = 4, \ldots, 10$ and

$$\frac{\partial P^j}{\partial PR_{t-\ell}} = \gamma\frac{\partial P^j}{\partial M3} + \frac{3\gamma}{10}\frac{\partial P^j}{\partial M10} \tag{2.50}$$

for $\ell = 1, 2, 3$. Equations 2.49 and 2.50 thus assume uniform changes in only years four to ten or one to three, respectively. Similar results hold for the standard-deviation variables, though here the change would be in $S_{t-\ell}$, the standard deviation of average daily precipitation across weeks. The magnitudes and signs of these partial effects can give an indication as to how much weight is given to either a short, recent past or a longer, more distant past in forming precipitation expectations and whether they increase or decrease the probability of observing a particular land use in the coming year. As such, they can provide insights as to how the occurrence of an extreme event, a severe drought for example, impact expectations.

Note, however, that the partial effects given in equations 2.47 and 2.48 only apply to hypothetical adjustments to precipitation in the past. As such, they can be viewed as how the probability of a particular field being devoted to a particular land use would have changed, had the precipitation in year $t - \ell$ been (marginally) different. Moving forward in time, however, equations 2.47 and 2.48 likely do not hold. To see this, consider time periods indexed by $t = 1, \ldots, T$. In year $t = 11$, $M3 = \frac{1}{3}\sum_{\ell=1}^{3} PR_{11-\ell}$ and $M10 = \frac{1}{10}\sum_{\ell=1}^{10} PR_{11-\ell}$. Then, in year $t = 12$, $M3 = \frac{1}{3}\sum_{\ell=1}^{3} PR_{12-\ell}$ and $M10 = \frac{1}{10}\sum_{\ell=1}^{10} PR_{12-\ell}$. Then, the changes in $M3$ and $M10$ can be calculated as:

$$\Delta M3 = M3_{t=12} - M3_{t=11} = \frac{1}{3}\left(PR_{11} - PR_8\right) \tag{2.51}$$

and

$$\Delta M10 = M10_{t=12} - M10_{t=11} = \frac{1}{10}\left(PR_{11} - PR_1\right). \tag{2.52}$$

In general, for a move from year t to $t+1$, these changes can be represented as:

$$\Delta M3 = M3_{t+1} - M3_t = \frac{1}{3}\left(PR_t - PR_{t-3}\right) \tag{2.53}$$

and

$$\Delta M10 = M10_{t+1} - M10_t = \frac{1}{10}\left(PR_t - PR_{t-10}\right). \tag{2.54}$$

The total change in probability moving from year t to $t+1$, $\Delta P^j = P^j_{t+1} - P^j_t$, can then be estimated as:

$$\Delta P^j = \frac{\partial P^j}{\partial M3}\Delta M3 + \frac{\partial P^j}{\partial M10}\Delta M10 \tag{2.55}$$

or

$$\Delta P^j = \frac{1}{3}\frac{\partial P^j}{\partial M3}\left(PR_t - PR_{t-3}\right) + \frac{1}{10}\frac{\partial P^j}{\partial M10}\left(PR_t - PR_{t-10}\right). \tag{2.56}$$

Similar results follow for the average-standard deviation variables, $S3$ and $S10$.

Clearly, there are many ways in which to calculate and thus interpret changes in the weather variables. Perhaps the most interesting are the dynamics implied by equations 2.54 and 2.55. However, these still likely do not fully capture the underlying dynamics due to the implication that there is no change in $S3$ and $S10$ given changes in $M3$ and $M10$, and vice versa. To fully account for all of the dynamics and simulate potential outcomes is outside the scope of the current study, but is recognized as an important area for future research. Thus, the results and interpretations offered in the subsequent section make use of simpler interpretations offered by equations 2.42, 2.49, and 2.50.

With respect to the standard partial effects given by equation 2.42, the study assumes an alternate-climate history wherein $M3$ and $M10$ or $S3$ and $S10$ have each increased by a common unit. We can thus get estimates of the average change in the probability of observing a particular land use via differentiation of P^j with respect to $M3$ and $M10$:

$$\Delta P^j = P^j_1 - P^j_0 = \gamma\frac{\partial P^j}{\partial M3} + \gamma\frac{\partial P^j}{\partial M10} \tag{2.57}$$

or $S3$ and $S10$:

$$\Delta P^j = P_1^j - P_0^j = \gamma \frac{\partial P^j}{\partial S3} + \gamma \frac{\partial P^j}{\partial S10} \tag{2.58}$$

where P_0^j denotes the probability under the actual weather history, P_1^j denotes the probability under the alternative weather history, and γ is the uniform increase. Results based on equations 2.57 and 2.58 can be thought of as a "naive forecast" of future impacts under a particular scenario.

Additionally, equations 2.49 and 2.50 and their $S3$ and $S10$ counterparts are used to examine the implications of alternative-weather histories where a measure is adjusted by γ for $t - \ell$ where $\ell = 1, 2, 3$ or $\ell = 4, \ldots, 10$. Note that summing equations 2.49 and 2.50 results in equation 2.57, and thus they represent a decomposition of equation 2.57 into two effects. If the past ten years are relevant in the formation of weather expectations, these two effects provide an indication as to whether the recent history or the more distant past plays a greater role in a farmer's assessment of the longer trend. If the short-term impact outweighs the long-term, this suggests that farmers may adjust there expectations quickly, which could lead to more erratic short-run land-use trends. Conversely, if the long-term impact is larger, producers may adjust their expectations more gradually, resulting in more gradual land-use-decision trends. The implications may be non-trivial. If farmer expectations are influenced more by longer trends — i.e., "signals" — they may be more likely to undertake appropriate land-use adaptations, albeit at a slower pace. If farmers place greater importance on the recent past, then adaptations may occur quickly, but are likely more prone to be inefficient adaptations or maladaptations if recent weather has deviated widely from the true underlying trend. In other words, if farmers are responding more to the "noise," they may make less-than-optimal decisions.

2.6.2 Average Transition Probabilities

Another interesting result that can be obtained from the dynamic specification is a matrix of transition probabilities. Because the lagged-dependent variables represent multiple scenarios (i.e., $\sum_{j=1}^{J} d_{i,t-1}^j = 1$), the standard approach to calculating APEs for the lagged-binary-dependent variables via discrete differences is not appropriate. This study estimates an average-transition probability matrix that provides the average-conditional probabilities of a field being allocated to land-use j in year t given the field was allocated to land-use k in year $t-1$. These values were calculated as:

$$\bar{P}\left(d_t^j = 1 \mid d_{t-1}^k = 1\right) = \frac{1}{N}\sum_{i=1}^{N} P_i\left(d_{t,i}^j = 1 \mid d_{t-1,i}^k = 1\right) \text{ for } j,k = 0,\ldots,4. \tag{2.59}$$

2.7 Results

2.7.1 Empirical Estimation and Limitations

Empirical analysis was carried out in LIMDEP 10. Estimation of the DML-RE proved to be difficult. Some difficulties resulted from the software while others appear to be data driven. With respect to LIMDEP 10, one limitation was that the multinomial logit with random effects is limited to 150 parameters. In many applications this likely does not present a problem. However, with five outcome categories (which realistically could be further divided), the number of included variables could be constrained. In the preferred model, there were 23 variables (including the constant) per equation, for a total of 92 parameters[3]. Due to data-driven difficulties, the 150-parameter constraint was not binding for this study, but it could be in similar studies with different data. For example, in random effects models, it is generally recommended that a full set of time-specific dummy variables be included (Wooldridge, 2010). Additionally, the model should include either the means or the full history of the time-varying explanatory variables, as is suggested by the framework from Wooldridge (2005). Thus, the number of variables can add up quickly, which may present issues in future studies.

Another software-imposed limitation worth mentioning is time. The random effects model for this study took considerable time to estimate, roughly two days. This was not altogether surprising given the nature of the problem, and may in fact be quicker than other software

[3]Recall that the parameters of the base case are normalized to zero.

packages. However, additional time burdens were imposed through what, for lack of better understanding, is deemed software instability. Particularly, this became an issue when estimating APEs. Often, the software would estimate the APEs for one or two categories and then exit, which would necessitate running the entire model again to get the remaining APEs (i.e., waiting a day or two). All of this is not to suggest, however, that LIMDEP not be used. In fact, it appears to be the most straightforward and flexible package with a built-in routine for a multinomial logit with random effects. While it may be possible to estimate the model in with other software packages, they may also encounter estimation difficulties or may not be as flexible[4]. LIMDEP also provides APEs (through the *PARTIALS* command) and can be prompted to give individual specific estimates of the random effects.

When what were assumed to be data-driven issues manifested, they typically resulted in a singular covariance matrix and subsequently no parameter estimates were produced. This limited the set of feasible models. One set which was subsequently excluded were the means of time-varying explanatory variables, e.g. mean prices, which would have been included to better adhere to the DML-RE framework. Inclusion of time dummies also resulted in this error[5]. Other explanatory variables that were subsequently excluded for the same reason include *ACRES* (the size of the field); *WEI*, a wind erodibility index; *PFERT*, *PLABOR*, and *PDIESEL* — input price indices for fertilizer, labor, and diesel. Lastly, this error prevented the use of correlated random effects in the final model presented in the results discussed below. Some of these results are likely the product of complete- or quasi-separation. These issues can arise when a particular variable or subset of variables perfectly, or nearly perfectly, predict the outcome, e.g., if, in the data, $COR_0 = 1$ and $SOY_{t-1} = 1$ is always associated with $COR_t = 1$.

[4]Other potential options may be the gsem command in Stata or the nlmefit function in MATLAB, for example. However, these may require substantially more knowledge of the functions or the dynamic-multinomial logit with random effects (DML-RE) theory. They likely will also place a heavier coding burden on the researcher, particularly with respect to estimating APEs.

[5]In some specifications, the inclusion of a time dummies or a time trend resulted in a crash following convergence, which could also be attributed to "software instability."

2.7.2 Model Estimates

Multiple specifications were tested for the DML-RE, but many were ruled out due to the estimation difficulties described above. In the preferred model, the following variables are included: *CLAY*, *SILT*, *PRECSUM*, *PCORN*, *PSOY*, *PWHEAT*, $M3$, $M10$, $S3$, $S10$, the initial conditions and lagged-dependent variables; a trend variable (*TREND*), and squared price terms. The price of sorghum, *PSOR* was dropped to aid with estimation. However, given the high correlation between sorghum and corn prices of 0.93 — due to the way they were calculated[6] — this is not believed to be a major concern. As such, *PCORN* can be viewed as a proxy for *PSOR*. Squared-price terms were included to capture potential non-linearities in the profit function with respect to prices. Other combinations to account for non-linearities were tried, such as interactions between prices and weather or prices and acres, but these too were met with estimation difficulties. The "other" category was chosen as the base case. Estimated parameter values for the corn, sorghum, soybeans, and wheat equations can be found in table 2.6 and additional estimation results are in table 2.7.

Land-use Transition Probabilities

The estimated transition probability matrix, shown in table 2.8, provides the average probability a field is devoted to land-use j in year t given that it was in land-use k in year $t-1$. Though there are many potential patterns that can be inferred from the table, some are more obvious. One example is continuous wheat. We can see evidence for this by looking across the $Wheat_{t-1}$ row and noting that the largest value is 0.54, associated with $Wheat_t$. This value suggests that, on average, if the previous land use for a field was wheat, then the land use in the current year will be wheat with a probability of 54%. As another example, we see that the average probability a field goes into sorghum in year t is, 25%, when wheat was planted in year $t-1$. While this is smaller than the probability of going back into wheat, it is the second largest probability given that a field was in wheat the year prior. Then, following sorghum, the largest average probability for the current year's land use is "other", at

[6]The *PSOR* variable was calculated using the same techniques described in section 2.5.2 but using corn futures as a proxy for sorghum futures, for which data were not available.

50%. Finally, following "other," the largest probability for the current year is wheat, at 86%. Recalling that the "other" category is largely comprised of fallow, this example is consistent with a wheat-sorghum-fallow rotation. The table also shows that the probability of going into corn is largest at 3%, when the previous crop was soybeans, and the largest probability of going into soybeans, 17%, is when the past crop was either corn or sorghum. Given that current land-use is either corn, soybeans, wheat, or other; the most likely outcome in the next period is wheat. If the current land use is sorghum, the most likely outcome in the next period is "other".

Marginal effects can be obtained by looking at differences in values down columns. For example, if land use in year $t-1$ was soybeans, the probability of observing soybeans decreases by about 12% compared to if land use in $t-1$ was corn or sorghum. Conversely, the probability of observing soybeans given soybeans in $t-1$ increases by 3% and 4%, respectively, compared to scenarios where wheat or "other" were observed in $t-1$. In general, the results presented in table 2.8 reinforce the idea that rotations are an important component in management decisions.

Price Effects

All APEs associated with crop prices, presented in table 2.9, were significant at the 1% level, except for the effect of $PWHEAT$ on soybeans, which was significant at the 5% level. An unexpected outcome was the negative APE associated with $PCORN$ on the probability of observing corn. However, the impact was negligible at about -0.0006. Corn represents a small share of production in the region, and may be grown primarily as an on-farm feed source, which could explain its non-responsiveness to price. Additionally, if other prices are rising along with corn prices, farmers growing corn for feed may opt to purchase cheaper feed and allocate the field to a more viable cash crop. The probabilities of observing soybeans or wheat were both positively affected by own-price increases, with APEs of 0.01 and 0.02 respectively. $PCORN$ — the sorghum-price proxy — had a positive impact on the probability of observing sorghum. The probability a field is allocated to "other" increases with $PCORN$

and decreases with $PSOY$ and $PWHEAT$. The direction of cross-price effects varied; see table 2.9. Across all crop-price effects, the magnitudes of the effects were small, particularly for corn. The largest positive cross-price effect was the impact of $PCORN$ on the "other" category, where it is estimated that a \$1 increase in $PCORN$ will result in a 4% increase in the probability a field is allocated to "other." $PCORN$ was also associated with the most negative price-effect: a \$1 increase in $PCORN$ decreases the probability of planting wheat by about 7%.

Weather Effects

Estimated APEs for the precipitation variables, shown in table 2.9, indicate that field-level decisions are influenced by past patterns and current planting conditions. In fact, many of the largest APEs (in absolute magnitude) are attributed to the weather variables. Precipitation APEs are all statistically significant at the 1% level except for two, which were not significant at the 1, 5, or 10% level.

The APEs associated with $PRECSUM$ are statistically significant at the 1% level for all categories. Results indicate that increased rainfall over the months of April to June decreases the probability that corn, sorghum, or "other" are observed and increases the probability of observing soybeans or wheat. The negative affects on corn and sorghum may be due to prevented plantings from continually wet fields, possibly with transitions to soybeans. The impact on wheat is more difficult to explain, as the precipitation on which $PRECSUM$ is based has not occurred at the time wheat is planted.

In general, the APEs on $M3$ and $M10$ suggest that long term increases (decreases) in precipitation will, ceteris paribus, lead to an increased (decreased) probability that a field is planted to soybeans and wheat and decrease (increase) the probabilities associated with the other three categories. This is similar to the results from Long et al. (2006), who estimated favorable outcomes for wheat and soybean yields under climate change scenarios. However, the results from Long et al. (2006) are due to changes in CO_2 levels, whereas the results here are due to increased precipitation.

To get an idea about the implications of long-term changes, we imagine an "alternate-history" scenario where $M3$ and $M10$ have both increased by 0.15 mm (i.e., $\gamma = 0.15$ in equations 2.57 and 2.58), which equates to roughly 1.08 additional inches over the six months[7]. The estimated average-change in the probabilities under this scenario are presented in table 2.10. The largest change in the region given the increase in total precipitation is for the probability of planting wheat at 8.94%. Soybean probabilities also would have increased, but less so at 0.95%. The largest decrease, -6.09%, is for the probability of observing the "other" land use. This result is reasonable when recalling that this category is dominated by fallow, a practice which is used to conserve water moisture. Under a scenario where there is greater precipitation, farmers likely see less of a need to fallow a field, thus decreasing the probability for this category. Potentially, the estimated increase in the wheat probability is due to wheat replacing fallow. Results also suggest a decrease in the corn and sorghum probabilities. The impact on corn probabilities is minimal, -0.12%, and thus likely results in little change in planted acreages. The results for $P(Corn)$ may be explainable through the same reasonings as the own-price effects. The impact on sorghum probabilities can be viewed in another way: if $M3$ and $M10$ had been 0.15 mm less, the sorghum probabilities would have been approximately 3.69% higher. In other words, sorghum, a drought tolerant crop, would have been more likely.

With respect to $S3$ and $S10$, results suggest that if weekly precipitation had been more variable, the probability of observing corn, soybeans, and wheat all decrease. The largest decrease is seen in the wheat probability at -2.57%; the corn and soybean impacts are minimal at -0.03% and -0.15% respectively. The probability a field is planted to sorghum shows a slight increase of 0.49%. This is consistent with the conclusion from Staggenborg et al. (2008) that sorghum may be preferable in areas with erratic rainfall. Additionally, if sorghum yields are more stable under stressful environments, this would be consistent with the suggestion from Olesen and Bindi (2002). A larger increase of 2.26% is seen with respect to the "other" category. This result may also be reasonable considering the dominance of

[7] Increase in inches calculated as $183 \times 0.15 \times 0.0393701$ where 183 is total number of days, 0.15 is the increase in average precipitation per day, and 0.0393701 is a millimeters to inches conversion factor.

fallow in this category. If increasingly variable weather patterns leads to greater uncertainty about receiving rainfall at crucial times, farmers may see greater benefits from a fallow period to conserve moisture for the next cash crop.

In all cases, the total impacts given in table 2.10 are dominated by the $M10$ or $S10$ APEs. This easily seen by looking at the (absolute) magnitudes of the 10-year variables relative to their three-year counterparts. This suggests that a longer view of weather patterns is more important in land-use decisions. The question then becomes what is driving their perception of this long term trend, the recent or more distant past? To get insights regarding the answer, we turn to equations 2.49 and 2.50, which decompose the total impacts from the alternate history described above into short- and long-term components. These decompositions are presented in table 2.11. The results largely indicate that the distant past — four to ten years prior — is the primary factor in how farmers perceive the long-term trend. This is seen by noting the absolute magnitudes of the components listed in table 2.11: the impacts of the long-term components are generally at least twice as large as their short-term counterparts. A notable exception is seen with respect to changes in soybean probabilities due to the alternate history for $M3$ and $M10$. In this case, the contribution from years one through three is estimated at 0.46% while that from years four through ten is 0.49%. Given the potentially high relative returns to soybeans, an interesting question for future research is whether economic incentives can cause farmers to be more willing to try and capitalize on short-term weather fluctuations. An alternative, but related, question could be whether the potential for greater profits actually impacts the process by which farmers form weather expectations. Intuitively, the answer would seem to be no: farmers form one expectation for weather and then make decisions accordingly. However, if the reverse is true, this would create interesting and potentially significant impacts for future research.

APE results suggest that farmers are in fact responding to weather patterns. Moreover, they suggest that long-term trends may be more important than short-term fluctuations. This would imply that, with respect to the question posed by Bradshaw et al. (2004), farmers are paying attention to the signals, and not just the noise. The results here have two implications. First, since farmers are responding to climate trends, the economics of the situation may induce adaptation. However, outside of land-use decisions, it is not possible to say from this study what other potential climate-change responses may be occurring. The second implication is with regards to the speed of adaptation. It appears that longer trends may be more important in terms of making land-use decisions, at least up to 10 years. Thus, even if farmers are adapting to climate patterns, it may not be at the speed that some would hope. However, responding to the signal, rather than the noise, should reduce the risk of long-term maladaptation.

It should be kept in mind that the results of this study are dependent on the region's characteristics, which is one of lower rainfall relative to other crop producing regions. As such, the results here may not be indicative of what should be expected elsewhere. Almost certainly, the responsiveness to changes in precipitation patterns will vary given the level of precipitation. Some regions, for example, may need only a slight increase in average precipitation levels to make corn or other water intensive crops economically feasible. As such, additional case studies are needed to better understand the total impacts of any climate change scenarios.

2.8 Conclusions

Climate change has been the subject of much research, primarily as it relates to food security, but also with respect to adaptation or mitigation measures that farmers can take in response to changing climates. Little, however, has been done to examine to which, if any, climate patterns farmers are responding. Because the issue of climate change is met with skepticism by some, there may be segments of the farming population that do not believe there is any need or are less willing to take adaptive measures, as found by Arbuckle, Morton, and Hobbs (2013) and Arbuckle, Prokopy, et al. (2013). This has potential ramifications for any study seeking to examine food security or other issues under climate change scenarios: the speed at which farmers do or do not respond will have consequences on food supplies, environmental impacts, etc. This study provides a framework for assessing the speed of land-use adaptation. Using this framework, the study examined the influence of multiple precipitation variables on field-level land-use decisions for 11 counties in central Kansas. The study used multiple precipitation variables on three- and ten-year histories to gain insights into what shapes producer weather expectations and how this may ultimately affect the speed of climate change adaptation.

Empirical analysis was done using a dynamic-multinomial logit with random effects (DML-RE) approach based on the dynamic multinomial probit with random effects model from Wooldridge (2005). This study appears to be the first to employ the DML-RE model for modeling farmer land-use decisions, which offers a better empirical approach as the presence of unobserved heterogeneity combined with lagged-dependent variables — the initial conditions problem — can play a crucial role in determining the entire path of outcomes and renders more traditional estimators inconsistent (W. H. Greene, 2012). Despite its theoretical appeal, however, the DML-RE presented estimation difficulties in this study. Potentially, this resulted from the need to include the initial conditions — binary variables — in addition to the lagged-dependent binary variables. This may create problems of complete- or quasi-separation in studies such as this one where there are multiple outcomes, and thus multiple binary variables, some of which have low rates of observance.

Average partial effects (APEs) from the estimated model indicate that both short- (three-year) and long-term (ten-year) precipitation histories factor into field-level land-use decisions in the region. This suggests that while farmers take a more long-term view of weather, their views are not so rigid that it inhibits the ability to change or respond in the short run. Simple APE results from indicate that had three- and ten-year average daily precipitation per week over the months of April-September been 0.15 millimeter higher — equivalent to a ~1.08 inch increase in total precipitation over these months — the region likely would have seen an increased share of fields devoted to wheat and soybeans and a smaller share to corn, sorghum, and "other," though the impact on corn is minimal. However, had the weather been characterized by higher variability, the results indicate the region would have seen a decrease in the share of fields planted to wheat or soybeans and an increase in the share devoted to sorghum or "other." Undoubtedly, there are interaction effects here that need exploration. For example, if precipitation becomes more variable but with respect to a higher mean or skewed towards higher precipitation, there are likely to be different outcomes in the land-use trends. Generally though, the results indicate that farmers are adapting to their local climates. Moreover, they indicate that farmer adaptations are largely driven by the longer historical patterns — ten years versus three years in this study — and are thus responding more to the "signal" than the "noise".

Several important implications may arise from the results of this study. The most obvious is the implication for global food security research. How, and at what speed, farmers adapt land-use decisions to a changing climate will play a key role in determining long-term food supply trends. However, it is unclear whether the results from this study would result in favorable or unfavorable outcomes for global food security, for a couple of reasons. First, this study represents just 11 Kansas counties, and so the results here likely can not be extrapolated to a global farming population. Additionally, the counties are not a major corn producing region, which is an important crop with respect to food security. Second, it is unclear what an optimal adaptation process would look like. For example, suppose farmers have perfect knowledge regarding future weather, which would lead to quicker adaptation and thus quicker changes in land-use trends. In this region, that may result in quicker transitions to trends characterized by more sorghum and fallow and less wheat and soybeans, assuming a future of lower rainfall. Absent this foresight, there may be a period of time during which wheat and soybeans replace fields that would have been allocated to sorghum or fallow with full foresight. If, hypothetically and clearly unrealistically, the rest of the world were characterized by the results of this study, this has important consequences. If having wheat and soybeans in place of sorghum and fallow is better for global food security, than the gradual process by which farmers adapt could actually have short-term benefits. Without knowing what optimal land-use trends are with respect to global food security, it is difficult to say how farmer adaptations will impact these scenarios. However, there could be a trade off between adaptation for the sake of farm profits and adaptation for the sake of food security. Additional research is needed in order to say anything definitive.

A second implication could be the impacts on or from crop insurance. It seems feasible that crop insurance may dampen the economic incentives to respond to a changing climate, at least for a period of time. Crop insurance could have a chronological-shifting effect, wherein the adaptation process follows a similar trend with or without crop insurance, but, with insurance, the process is shifted down the time line. The process by which farmers adapt could also impact crop insurance indemnity payments. For the region studied here, this seems most likely in the case of soybeans, where the perceptions of the long-term trend appear to be heavily influenced by the past three years. This could potentially result in false optimism for the upcoming season and thus a decision to grow soybeans. If the previous three years represent a favorable departure from the long-term trend, this could lead to unfavorable profit outcomes. The impacts on crop insurance, however, will depend on the structure of the program and thus warrants further research.

As a final implication example for this study, we consider the impacts on or from irrigation and water conservation. As with crop insurance, it seems likely that the ability to irrigate a field would impact how and at what speed farmers adapt to a changing climate. On the other hand, in areas where irrigation is possible — e.g., portions of western Kansas which rely on the Ogallala aquifer — if future weather is characterized by lower rainfall, this could potentially increase irrigation rates and thus lead to quicker depletion of the aquifer. Actual outcomes will depend on actual weather trends and also on state or local laws, such as water rights, irrigation limits, etc. Obviously, forcing adaptation by dictating land uses — e.g., sorghum instead of corn — is not politically feasible or desirable, but incentivizing it may be. Policies to promote water conservation practices or increased irrigation efficiency, to the extent that this has not already been done, may also be warranted. Again, additional research is needed in this area to draw any strong conclusions.

Figures

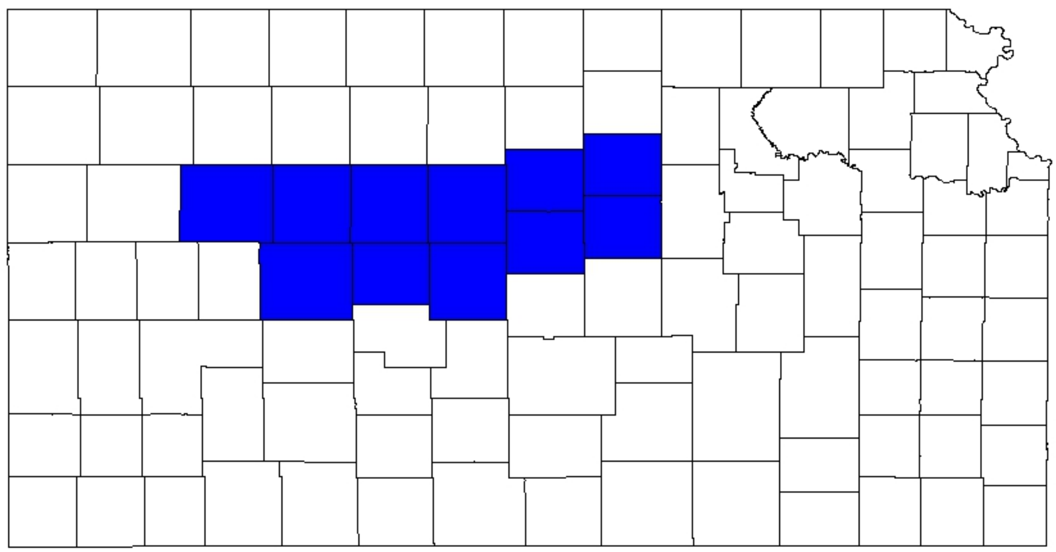

Figure 2.1: Central-Western Kansas study area

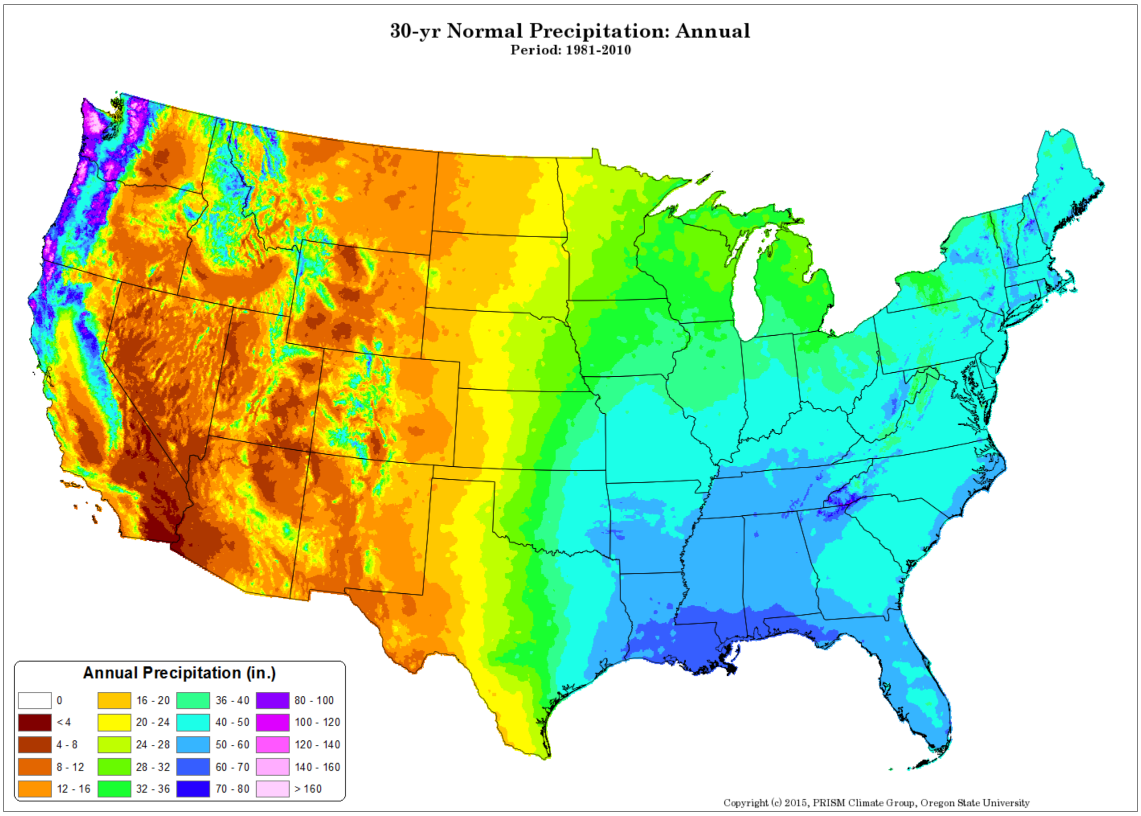

Figure 2.2: 30-Year normal precipitation 1981-2010

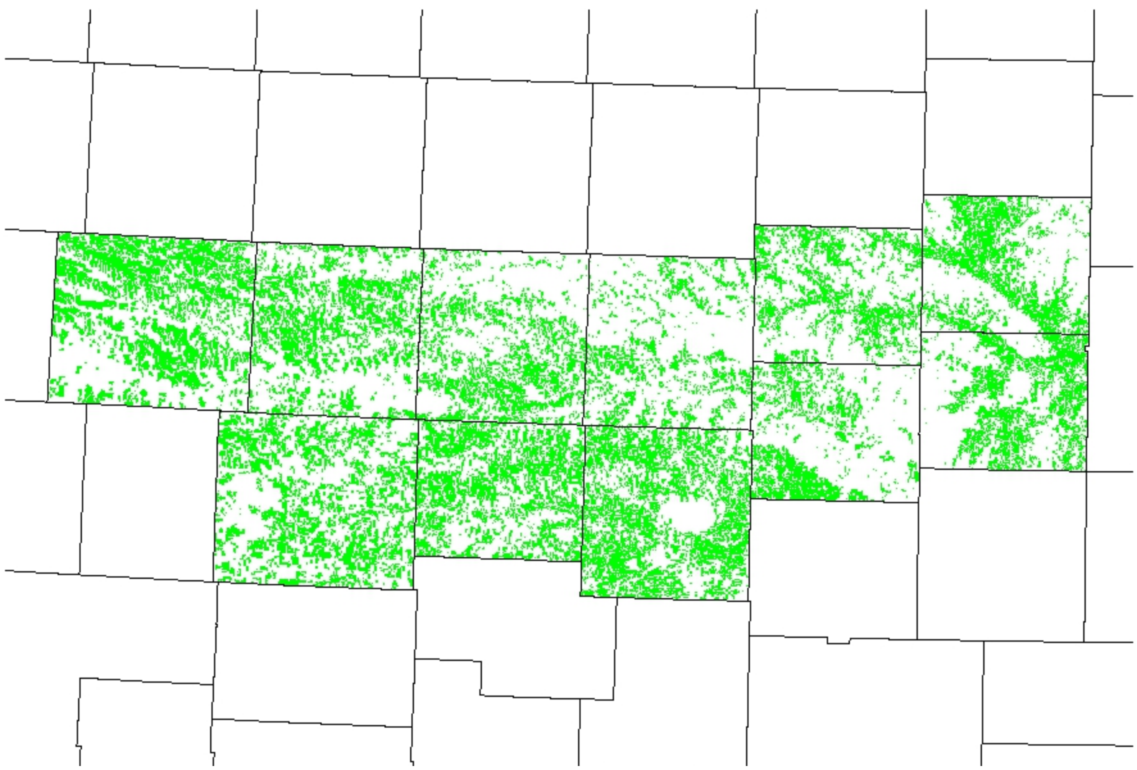

Figure 2.3: Fields used in analysis

Tables

Table 2.1: *Share of fields by land-use category and year*

Year	Corn	Sorghum	Soybeans	Wheat	Other
2003	0.02	0.12	0.03	0.62	0.21
2004	0.03	0.19	0.07	0.52	0.18
2005	0.04	0.08	0.08	0.65	0.15
2006	0.03	0.15	0.03	0.56	0.24
2007	0.03	0.17	0.02	0.58	0.19
2008	0.04	0.20	0.04	0.55	0.18
2009	0.05	0.20	0.06	0.48	0.21
2010	0.06	0.17	0.08	0.49	0.20
2011	0.06	0.19	0.08	0.46	0.21
2012	0.10	0.19	0.07	0.45	0.21
All	0.05	0.17	0.06	0.53	0.20

Table 2.2: *Share of acres by land-use category and year*

Year	Corn	Sorghum	Soybeans	Wheat	Other
2003	0.03	0.14	0.02	0.59	0.22
2004	0.04	0.18	0.06	0.51	0.22
2005	0.05	0.09	0.06	0.63	0.17
2006	0.04	0.15	0.03	0.53	0.25
2007	0.05	0.17	0.02	0.55	0.21
2008	0.05	0.19	0.04	0.52	0.20
2009	0.07	0.19	0.05	0.46	0.23
2010	0.09	0.16	0.08	0.46	0.21
2011	0.08	0.18	0.07	0.44	0.23
2012	0.13	0.18	0.06	0.41	0.22
All	0.06	0.16	0.05	0.51	0.22

Table 2.3: *Descriptions and shares for land-use explanatory variables*

Variable	Description	Average
Lagged-dependent variables		
COR_{t-1}	Binary, = 1 if land use in period $t = 0$ was corn	0.04
SOR_{t-1}	Binary, = 1 if land use in period $t = 0$ was sorghum	0.17
SOY_{t-1}	Binary, = 1 if land use in period $t = 0$ was soybeans	0.03
WHT_{t-1}	Binary, = 1 if land use in period $t = 0$ was wheat	0.55
OTH_{t-1}	Binary, = 1 if land use in period $t = 0$ was other	0.21
Initial conditions variables		
COR_0	Binary, = 1 if land use in period $t = 0$ was corn	0.02
SOR_0	Binary, = 1 if land use in period $t = 0$ was sorghum	0.12
SOY_0	Binary, = 1 if land use in period $t = 0$ was soybeans	0.03
WHT_0	Binary, = 1 if land use in period $t = 0$ was wheat	0.62
OTH_0	Binary, = 1 if land use in period $t = 0$ was other	0.21

Table 2.4: *Descriptions and summary statistics for non-land-use explanatory variables*

Variable	Description	Mean	Std.Dev.	Minimum	Maximum
CLAY	Percent clay in soil	0.28	0.04	0.05	0.54
SILT	Percent silt in soil	0.56	0.08	0.05	0.70
PCORN	Expected price of corn	3.96	1.30	2.07	6.13
PSOY	Expected price of soybeans	8.75	2.61	5.31	12.90
PWHEAT	Expected price of wheat	5.03	1.64	2.99	7.64
PRECSUM	Total precipitation (mm) Apr-Jun	2.63	0.81	0.95	6.94
M_3	Weekly precipitation (mm) Apr-Sep, 3-year average	2.53	0.40	1.64	3.73
M_{10}	Weekly precipitation (mm) Apr-Sep, 10-year average	2.55	0.26	1.81	3.23
S_3	Std. dev. of weekly precipitation Apr-Sep, 3-year average	2.81	0.36	1.71	4.14
S_{10}	Std. dev. of weekly precipitation Apr-Sep, 10-year average	2.89	0.25	1.99	3.64

Table 2.5: *Yield vs. evapotranspiration relationship for crops of the central High Plains*

Crop	Max ET for full season variety	Threshold ET	Slope of yield vs. ET
Corn	25 in.	10.9 in.	16.9 bu./ac./in.
Grain Sorghum	21 in.	6.9 in.	12.2 bu./ac./in.
Soybeans	24 in.	7.8 in.	4.6 bu./ac./in.
Wheat	24 in.	10.0 in.	6.0 bu./ac./in.

Table 2.6: *Random-effects dynamic-multinomial logit parameter estimates*

Variable	$P\,(Corn)$	$P\,(Sorghum)$	$P\,(Soybeans)$	$P\,(Wheat)$
$TREND$	0.14^{***} (13.15)	-0.04^{***} (-6.38)	-0.15^{***} (-12.88)	0.01 (0.94)
$CLAY$	-0.13 (-0.49)	-0.84^{***} (-5.89)	0.26 (1.24)	0.52^{***} (4.32)
$SILT$	-3.77^{***} (-30.99)	0.30^{***} (4.31)	-1.10^{***} (-11.16)	-0.33^{***} (-5.64)
$PRECSUM$	0.00 (0.15)	-0.03^{***} (-2.79)	0.19^{***} (10.55)	0.13^{***} (13.1)
$PCORN$	-2.12^{***} (-13.24)	0.32^{***} (3.73)	-2.66^{***} (-17.78)	-1.33^{***} (-17.49)
$PSOY$	1.22^{***} (14.26)	0.51^{***} (10.63)	1.85^{***} (20.93)	0.39^{***} (9.14)
$PWHEAT$	-1.56^{***} (-11.85)	0.07 (1.01)	-0.98^{***} (-7.41)	1.57^{***} (23.07)
$PCORN^2$	0.20^{***} (11.9)	-0.03^{***} (-3.44)	0.15^{***} (8.79)	0.12^{***} (14.27)
$PSOY^2$	-0.06^{***} (-12.5)	-0.02^{***} (-9.47)	-0.06^{***} (-12.27)	-0.01^{***} (-6.06)
$PWHEAT^2$	0.15^{***} (12.6)	0.00 (-0.60)	0.11^{***} (9.36)	-0.14^{***} (-23.07)
$M3$	0.58^{***} (7.07)	0.20^{***} (4.33)	0.56^{***} (8.83)	-0.70^{***} (-18.39)
$M10$	-1.43^{***} (-11.09)	0.78^{***} (12.12)	5.68^{***} (62.69)	4.09^{***} (83.41)
$S3$	0.20^{***} (3.35)	-0.08^{**} (-2.17)	0.23^{***} (4.61)	0.35^{***} (11.87)
$S10$	-1.83^{***} (-20.7)	-0.65^{***} (-14.53)	-1.73^{***} (-26.33)	-1.54^{***} (-44.14)
COR_{t-1}	-0.72^{***} (-19.74)	0.27^{***} (8.01)	1.20^{***} (29.18)	-2.09^{***} (-80.12)
SOR_{t-1}	-0.47^{***} (-12.37)	-0.10^{***} (-4.65)	0.37^{***} (12.94)	-3.71^{***} (-240.1)
SOY_{t-1}	1.31^{***} (29.52)	0.76^{***} (22.96)	0.36^{***} (10.35)	-1.25^{***} (-50.7)
WHT_{t-1}	1.06^{***} (34.28)	1.65^{***} (89.26)	-0.40^{***} (-14.63)	-1.32^{***} (-133.01)
COR_0	5.08^{***} (113.23)	0.76^{***} (18.33)	1.95^{***} (42.29)	0.69^{***} (20.52)
SOR_0	0.91^{***} (28.11)	1.19^{***} (65.11)	0.65^{***} (22.65)	1.16^{***} (69.57)
SOY_0	1.82^{***} (33.73)	1.16^{***} (34.78)	1.49^{***} (38.06)	0.81^{***} (27.55)
WHT_0	-0.13^{***} (-5.15)	0.41^{***} (32.52)	0.42^{***} (18.9)	1.13^{***} (106.18)
α_j	6.23^{***} (18.29)	-5.05^{***} (-25.58)	-14.63^{***} (-44.51)	-6.81^{***} (-40.02)
σ_a	2.45^{***} (155.73)	0.77^{***} (152.49)	0.83^{***} (99.02)	0.88^{***} (213.9)

Values in parentheses denote z-statistics
***, **, * $\Rightarrow$ Significance at 1%, 5%, 10% level

Table 2.7: *Model estimation details*

Log-likelihood function	-417,422
N	409,878
Akaike Information Criterion (AIC)	835,036
AIC/N	2.037
Halton Draws	500

Table 2.8: *Average transition probability matrix*

	$Corn_t$	$Sorghum_t$	$Soybeans_t$	$Wheat_t$	$Other_t$
$Corn_{t-1}$	0.01	0.11	0.17	0.42	0.29
$Sorghum_{t-1}$	0.02	0.14	0.17	0.18	0.50
$Soybeans_{t-1}$	0.03	0.12	0.05	0.61	0.19
$Wheat_{t-1}$	0.02	0.25	0.02	0.54	0.16
$Other_{t-1}$	0.01	0.03	0.01	0.86	0.09

Table 2.9: *Selected average partial effects*

Variable	$P\,(Corn)$	$P\,(Sorghum)$	$P\,(Soybeans)$	$P\,(Wheat)$	$P\,(Other)$
$PCORN$	$-6.10E-04^{***}$ (-4.86)	0.05^{***} (16.54)	-0.02^{***} (-17.55)	-0.07^{***} (-15.39)	0.04^{***} (13.01)
$PWHEAT$	$-2.70E-04^{***}$ (-5.81)	-0.01^{***} (-6.9)	0.00^{**} (2.53)	0.02^{***} (11.97)	-0.01^{***} (-11.24)
$PSOY$	$2.60E-04^{***}$ (4.73)	-0.01^{***} (-4.55)	0.01^{***} (19.39)	0.01^{***} (6.52)	-0.02^{***} (-13.8)
$PRECSUM$	$-2.00E-04^{***}$ (-3.88)	-0.02^{***} (-14.31)	0.00^{***} (6.79)	0.03^{***} (17.21)	-0.01^{***} (-10.04)
$M3$	$2.44E-03^{***}$ (12.95)	0.10^{***} (20.84)	0.02^{***} (17.9)	-0.19^{***} (-28.5)	0.07^{***} (13.95)
$M10$	-0.01^{***} (-25.4)	-0.35^{***} (-50.3)	0.05^{***} (22.04)	0.78^{***} (85.96)	-0.47^{***} (-72.72)
$S3$	$-7.00E-05$ (-0.52)	-0.05^{***} (-13.06)	$1.70E-04$ (0.23)	0.09^{***} (16.47)	-0.04^{***} (-9.33)
$S10$	$-1.68E-03^{***}$ (-7.27)	0.08^{***} (14.91)	-0.01^{***} (-9.2)	-0.26^{***} (-35.99)	0.19^{***} (39.32)

Values in parentheses denote z-statistics

***, **, * $\Rightarrow$ Significance at 1%, 5%, 10% level

Table 2.10: *Calculated average change in probabilities under alternative weather history*

Scenario	$P\,(Corn)$	$P\,(Sorghum)$	$P\,(Soybeans)$	$P\,(Wheat)$	$P\,(Other)$
Change in $M3$ and $M10$	-0.12%	-3.69%	0.95%	8.94%	-6.09%
Change in $S3$ and $S10$	-0.03%	0.49%	-0.15%	-2.57%	2.26%

Changes calculated using increase of 0.15 mm in scenario variables.

Table 2.11: *Calculated average change in probabilities under alternative weather history decomposed into short- and long-term components*

Total Impact From	$P\,(Corn)$	$P\,(Sorghum)$	$P\,(Soybeans)$	$P\,(Wheat)$	$P\,(Other)$
	Change in $M3$ and $M10$				
Years 1 to 3	-0.01%	-0.05%	0.46%	0.72%	-1.12%
Years 4 to 10	-0.11%	-3.63%	0.49%	8.22%	-4.97%
	Change in $S3$ and $S10$				
Years 1 to 3	-0.01%	-0.37%	-0.04%	0.12%	0.29%
Years 4 to 10	-0.02%	0.85%	-0.10%	-2.70%	1.96%

Changes calculated using increase of 0.15 mm in weather variables.

Chapter 3

An artificial neural network approach to acreage-share modeling

3.1 Introduction

Agricultural land-use patterns can, in one instance, be impacted by exogenous shocks while in another may impose a shock elsewhere. Global or local market shocks — such as the Renewable Fuel Standard or an individual ethanol plant — can influence land-use patterns at either the macro or micro levels (see, e.g., Searchinger et al. (2008), Hertel et al. (2010), Plevin, O'Hare, Jones, Torn, and Gibbs (2010)). Conversely, land-use patterns can effect micro or macro markets through supply-shock impacts. Government policies can also influence land-use patterns, through direct price distortions, caused by subsidies or price floors for example. Non-price policies can also directly impact land use, such as the 1985 Farm Bill conservation provisions that intended, at least in part, to stem the conversion of highly erodible lands to crop production (Malone, 1986). Crop distributions will be influenced by environmental constraints as well, such as soil quality or water availability via rainfall or irrigation. Changes in land-use patterns can in turn affect the local environment through subsequent changes in sediment or nutrient runoff. The bilateral relationship between land-use and the rest of the political-economy, combined with the economic importance of agriculture and its connected industries for many individuals, place great importance on understanding the factors that influence land-use patterns.

Assuming land-use shares take a logistic form is an often-used approach to modeling land-use patterns (Wu & Segerson, 1995). This approach has been used, for example, in the analysis of groundwater pollution (Wu & Segerson, 1995); to examine the costs of carbon sequestration (Plantinga, Mauldin, & Miller, 1999); and to examine climate change adaptation by South American farmers (Seo & Mendelsohn, 2008). Chakir and Le Gallo (2013) note three key reasons for using this approach: it ensures predicted shares are strictly between zero and one; it is parsimonious in parameters; and empirical applications are simplified via a "log-linear transformation."

However, to quote Hirschman (1984, p. 11) (who was paraphrasing Sen (1977)), "...parsimony in theory construction can be overdone and something is sometimes to be gained by making things more complicated." A similar sentiment is echoed by Neal (1996, p. 103): "Sometimes a simple model may outperform a more complex model, at least when the training data is limited. Nevertheless, I believe that deliberately limiting the complexity of the model is not fruitful when the problem is evidently complex." While Hirschman (1984) was speaking to parsimony in economic theory and Neal (1996) was speaking to engineering applications of artificial neural networks (ANN), it is likely their views are relevant to modeling land use.

A primary motivation for this study is that the functional form underlying standard "linear-logit" acreage-response models may, in some situations, be more accurately modeled using an ANN approach. Literature regarding ANNs suggest that they are well suited for tasks where the true underlying function is unknown. Hornik, Stinchcombe, and White (1990), for example, show that a properly specified ANN is capable of approximating an arbitrary function $f(\mathbf{x}) : \mathbb{R}^k \to \mathbb{R}^\ell$ and its derivatives to an arbitrary level of accuracy. Hornik (1991) expands on the work of Hornik et al. (1990) and similar studies by showing that many of the explicit assumptions often employed to obtain the conclusions are unnecessary. Given very general conditions, Hornik et al. (1990, p. 252) conclude that "for arbitrary input environment measures μ, standard multilayer feed-forward networks... can approximate any function on $L^p(\mu)$ (the space of all functions on $\mathbb{R}^k$ such that $\int_{\mathbb{R}^k} |f(x)|^p \, d\mu(x) < \infty$) arbitrarily well if closeness is measured by $\rho_{p,\mu}$" where

$$\rho_{p,\mu}(f,g) = \left[\int_{\mathbb{R}^k} |f(x) - g(x)|^p \, d\mu(x)\right]^{1/p} \tag{3.1}$$

and $1 \leq p < \infty$. These results have two important implications for the current study. First, even if the traditional specification of land-use empirical models — one of a linear-in-parameters and linear-in-explanatory variables index function — represents the true underlying model, it can be approximated with an appropriate ANN. Second, if the traditional linear-index specification is incorrect, an ANN approach can allow the modeler to avoid inaccurate or inconsistent inferences, such as elasticity estimates, from functional-form misspecification. Additionally, because estimation of multiple outputs is done simultaneously with ANNs, this method may provide some of the same benefits as the traditional seemingly unrelated regression (SUR) approach. Specifically, it may account for the contemporaneous dependence between equations.

As implied, however, there is a trade off when switching to an ANN approach. While estimation of model parameters is generally not an issue — ANN estimation is available in many statistical software packages — moving past prediction and into inference or estimation of elasticities is likely to be more taxing to the researcher. There will also likely be an increased burden on computer resources. For ANNs to be seen as a viable alternative to traditional land-use modeling approaches, it must be shown that they produce reasonable, if not better results, with respect to measures such as model fit and elasticities[1].

The primary purpose of this study is to examine the viability of ANNs as an alternative to the traditional linear-logit land-use models. Using Kansas land use data as an empirical application, the ANN approach is compared with an extension of the linear-logit model as summarized by Wu and Brorsen (1995). The empirical framework for the models is

[1]For an ANN to be "better" in terms of model fit is relatively straightforward. However, with respect to elasticities, it is difficult to say which approach provides "better" results, as the true measures are unknown. Despite the results from Hornik et al. (1990) and Hornik (1991), it remains difficult to say whether or not the true underlying function and its derivatives have been captured. Thus, for elasticities, it is hoped merely that ANN estimates seem "reasonable", a distinction that is left to the reader.

provided in section 3.2. Section 3.3 outlines the empirical methods for both approaches as well as elasticity calculations (one basis of comparison) under each. Section 3.4 provides a description of the empirical application and associated data. In addition to estimated-elasticity comparisons, the two approaches are compared with respect to model fit. The results of these comparisons are presented in section 3.5. Concluding remarks and discussion are provided in 3.6.

3.2 Modeling Framework

This study assumes that a farmer seeks to maximize expected profit on a particular field by choosing between $j = 0, \ldots, J$ crops to which the field may be planted. Under this scenario, the field is planted to crop j if:

$$\Pi_j > \Pi_k \text{ for all } k \neq j, \tag{3.2}$$

where Π_j is the expected profits from crop j. Because Π_j is unobservable by the researcher, it is decomposed into an observable component, π_j, and an unobservable, stochastic component, ε_j. Thus, condition 3.2 can be rewritten as:

$$\pi_j + \varepsilon_j > \pi_k + \varepsilon_j \text{ for all } k \neq j. \tag{3.3}$$

It is assumed that the observable component of expected profits can be represented by:

$$\pi_j = g\left(\mathbf{x}_j; \boldsymbol{\beta}_j\right), \tag{3.4}$$

where $\mathbf{x}$ is a vector of explanatory variables and $\boldsymbol{\beta}$ is a vector of parameters. Then, using equation 3.4, condition 3.3 becomes:

$$g\left(\mathbf{x}_j; \boldsymbol{\beta}_j\right) + \varepsilon_j > g\left(\mathbf{x}_k; \boldsymbol{\beta}_k\right) + \varepsilon_k \text{ for all } k \neq j. \tag{3.5}$$

Decomposing the farmer's problem as in condition 3.5 allows it to be viewed from a probabilistic perspective. That is, the probability that crop j is planted, P_j, is:

$$P_j = P\left(g\left(\mathbf{x}_j; \boldsymbol{\beta}_j\right) + \varepsilon_j > g\left(\mathbf{x}_k; \boldsymbol{\beta}_k\right) + \varepsilon_k \text{ for all } k \neq j\right) \tag{3.6}$$

or

$$P_j = P\left(\varepsilon_k - \varepsilon_j < g\left(\mathbf{x}_j; \boldsymbol{\beta}_j\right) - g\left(\mathbf{x}_k; \boldsymbol{\beta}_k\right) \text{ for all } k \neq j\right). \tag{3.7}$$

It is typically assumed that the error terms, ε_j for $j = 0, \ldots, J$, are independently and identically distributed with a Gumbel distribution and that the share of land allocated to crop j, s_j, is equal to the probability that a given field is planted to crop j (Wu & Adams, 2003). Under these assumptions, the share of land devoted to crop j in a region is given by

$$s_j = P_j = \frac{\exp\left(g\left(\mathbf{x}_j; \boldsymbol{\beta}_j\right)\right)}{\sum_{j=0}^{J} \exp\left(g\left(\mathbf{x}_j; \boldsymbol{\beta}_j\right)\right)}. \tag{3.8}$$

The identity given in equation 3.8 provides the basis for the empirical procedures outlined in section 3.3.

3.3 Empirical Models

Two approaches are used to estimate the share equations given in equation 3.8. The first combines the cross-sectionally heteroskedastic and time-wise autoregressive model from Kmenta (1986) and the seemingly unrelated regression technique from Zellner (1962). This procedure — dubbed the SUR-HEAR model — was proposed by Wu and Brorsen (1995). The second approach is to estimate crop shares using artificial neural networks (ANNs).

With equation 3.8 as a starting point, two adjustments are made before coming to the point of divergence between the SUR-HEAR and ANN approaches. First, as noted by W. H. Greene (2012), an indeterminacy in the model must be removed. Because the $J + 1$ probabilities sum to one, the probabilities can be reproduced by using $\boldsymbol{\beta}_j^* = \boldsymbol{\beta}_j + \mathbf{q}$ for any vector $\mathbf{q}$ (W. H. Greene, 2012). This problem can be resolved by setting all parameters for the 0^{th} crop equal to zero; that is, setting $\boldsymbol{\beta}_0 = \mathbf{0}$ (W. H. Greene, 2012). Then, as long as $g(\mathbf{x}_0; \boldsymbol{\beta}_0) = 0$, equation 3.8 becomes

$$s_j = P_j = \frac{\exp\left(g\left(\mathbf{x}_j; \boldsymbol{\beta}_j\right)\right)}{1 + \sum_{j=1}^{J} \exp\left(g\left(\mathbf{x}_j; \boldsymbol{\beta}_j\right)\right)} \text{ for } j = 1, \ldots, J \tag{3.9}$$

and

$$s_0 = P_0 = \frac{1}{1 + \sum_{j=1}^{J} \exp\left(g\left(\mathbf{x}_j; \boldsymbol{\beta}_j\right)\right)} \text{ for } j = 0. \tag{3.10}$$

It will be shown in subsections 3.3.1 and 3.3.2 that $g(\mathbf{x}_0; \boldsymbol{\beta}_0) = 0$ if $\boldsymbol{\beta}_0 = \mathbf{0}$ for both the SUR-HEAR and ANN approaches.

Noting that equation 3.9 represents a highly nonlinear system, the second adjustment is made with the goal of simplifying estimation. The first step towards simplification is to divide each of the J equations in 3.9 by s_0 in equation 3.10:

$$\frac{s_j}{s_0} = \exp\left(g\left(\mathbf{x}_j; \boldsymbol{\beta}_j\right)\right) \text{ for } j = 1, \ldots, J. \tag{3.11}$$

Taking the natural logarithm of equation 3.11 leaves the estimable system

$$\ln\left(\frac{s_j}{s_0}\right) = g\left(\mathbf{x}_j; \boldsymbol{\beta}_j\right) \text{ for } j = 1, \ldots, J. \tag{3.12}$$

A key difference between traditional acreage-response models and the ANN approach employed here is in the specification of $g\left(\mathbf{x}_j; \boldsymbol{\beta}_j\right)$. Commonly, a linear function is chosen, such that $g\left(\mathbf{x}_j; \boldsymbol{\beta}_j\right) = \mathbf{x}_j' \boldsymbol{\beta}_j$. With the ANN approach, this linear form is replaced with a semi-nonparametric flexible functional form, as will be outlined in section 3.3.2 below.

3.3.1 SUR-HEAR Model

Given the contemporaneous correlation between regression residuals across crop-share equations, the SUR model is an obvious choice for estimation. However, the SUR model makes two assumptions which Wu and Brorsen (1995) posit are unlikely to hold in an acreage-response model. First, the SUR model assumes strict homoskedasticity (W. H. Greene, 2012). Letting $i = 1, \ldots, N$ denote regions and $t = 1, \ldots, T$ denote time periods, this assumption implies that

$$E\left(\boldsymbol{\varepsilon}_j \boldsymbol{\varepsilon}_j' \mid \mathbf{X}_1, \ldots, \mathbf{X}_J\right) = \sigma_{jj} \mathbf{I}_{NT} \tag{3.13}$$

where $\boldsymbol{\varepsilon}_j = \left[\varepsilon_{1,1,j} \ldots \varepsilon_{1,T,j} \ldots \varepsilon_{N,1,j} \ldots \varepsilon_{N,T,j}\right]'$ and NT denotes the total number of observations for each crop-share $j = 1, \ldots, J$. Wu and Brorsen (1995) suggest that condition 3.13 is unlikely to hold, given factors such as differing county sizes and cultivation histories, for example. This study also assumes that homoskedasticity across an equation is unlikely, though given the similarity in region sizes (see section 3.4), heteroskedasticity may enter through other channels, such as local policies or market factors such as proximity to ethanol production.

The second assumption the SUR model makes is that the error terms are correlated across equations but uncorrelated across observations (W. H. Greene, 2012):

$$E\left(\varepsilon_{i,t,j} \varepsilon_{m,s,k} \mid \mathbf{X}_1, \ldots, \mathbf{X}_J\right) = \begin{cases} \sigma_{j,k} & \text{for } t = s, i = m \\ 0 & \text{for } t \neq s \text{ or } i \neq m \end{cases}. \tag{3.14}$$

This assumption implies that autocorrelation is not present in the data. For land use shares, this may not hold due to the prevalence of crop rotations; multiple-year climate patterns (e.g. prolonged drought); or other factors.

The SUR-HEAR model proposed by Wu and Brorsen (1995) corrects for heteroskedasticity and autocorrelation via the cross-sectionally heteroskedastic and time-wise autoregressive (HEAR) model from Kmenta (1986). Assuming $g\left(\mathbf{x}_j; \boldsymbol{\beta}_j\right) = \mathbf{x}_j' \boldsymbol{\beta}_j$, the crop share for crop $j = 1, \ldots, J$ is given by

$$s_j = P_j = \frac{\exp\left(\mathbf{x}_j' \boldsymbol{\beta}_j\right)}{1 + \sum_{j=1}^{J} \exp\left(\mathbf{x}_j' \boldsymbol{\beta}_j\right)} \tag{3.15}$$

and the estimated equations become

$$\ln\left(\frac{s_j}{s_0}\right) = \mathbf{x}_j' \boldsymbol{\beta}_j. \tag{3.16}$$

Note that under this specification, $g(\mathbf{x}_0; \boldsymbol{\beta}_0) = \mathbf{x}_0' \boldsymbol{\beta}_0 = 0$ for $\boldsymbol{\beta}_0 = \mathbf{0}$ as required for the results given in equations 3.9 and 3.10.

The SUR-HEAR approach starts by estimating each equation given in equation 3.16 via ordinary least squares (OLS). Using the regression residuals, $\tilde{\varepsilon}_{i,t,j}$, the autocorrelation coefficient, $\rho_{i,j}$, is estimated as

$$\tilde{\rho}_{i,j} = \frac{\sum_{t=2}^{T} \tilde{\varepsilon}_{i,t,j} \tilde{\varepsilon}_{i,t-1,j}}{\sum_{t=1}^{T} \tilde{\varepsilon}_{i,t,j}^2}. \tag{3.17}$$

This estimate is then used to transform the data such that:

$$y_{i,1,j}^* = y_{i,1,j} \sqrt{1 - \tilde{\rho}_{i,j}^2}, \tag{3.18}$$

$$y_{i,t,j}^* = y_{i,t,j} - \tilde{\rho}_{i,j} y_{i,t-1,j} \text{ for } t = 2, \ldots, T, \tag{3.19}$$

$$x_{i,1,j}^* = x_{i,1,j} \sqrt{1 - \tilde{\rho}_{i,j}^2}, \tag{3.20}$$

and

$$x_{i,t,j}^* = x_{i,t,j} - \tilde{\rho}_{i,j} x_{i,t-1,j} \text{ for } t = 2, \ldots, T, \tag{3.21}$$

where $y_{i,t,j} = \ln\left(\frac{s_{i,t,j}}{s_{i,t,0}}\right)$.

The next step in the SUR-HEAR procedure is to correct for heteroskedasticity, and begins with again estimating the share equations in equation 3.16 via OLS using the transformed data $\mathbf{y}_j^*$ and $\mathbf{X}_j^*$ for $j = 1, \ldots, J$. From the estimated equations, a new set of regression residuals, $\tilde{\varepsilon}_{i,t,j}^*$, are obtained. These new residuals are then used to estimate a separate error variance for each crop in each region:

$$\hat{\sigma}_{i,j} = \sum_{t=1}^{T} \frac{\left(\tilde{\varepsilon}_{i,t,j}^*\right)^2}{T}. \tag{3.22}$$

Once $\hat{\sigma}_{i,j}$ is obtained, the data is transformed once again such that

$$y_{i,t,j}^{**} = \frac{y_{i,t,j}^*}{\hat{\sigma}_{i,j}} \tag{3.23}$$

and

$$x^{**}_{i,t,j} = \frac{x^{*}_{i,t,j}}{\hat{\sigma}_{i,j}}. \tag{3.24}$$

Finally, Wu and Brorsen (1995) apply the seemingly unrelated regression (SUR) estimator to the transformed equations given by

$$y^{**}_{i,t,j} = \mathbf{x}^{**\prime}_{i,t,j}\boldsymbol{\beta}_j + \varepsilon^{**}_{i,t,j}. \tag{3.25}$$

This last step is used to account for the contemporaneous correlation across the $j = 1, \ldots, J$ equations.

3.3.2 Artificial Neural Networks

Fausett (1994, p. 3) defines an artificial neural network (ANN) as "an information-processing system that has certain performance characteristics in common with biological neural networks." Thus, ANNs can be viewed as the parallel interconnection of many simple elements known as neurons (also referred to as nodes) (West, Brockett, & Golden, 1997). ANNs process information by passing signals between neurons along arcs, which are weighted according to the usefulness of the information being sent. As the network is estimated, weights are adjusted so that the useful arcs are strengthened until the network learns to recognize patterns in the data. The objective is to have the network learn these patterns in such a way that they can be generalized and used to classify new data (Fausett, 1994; West et al., 1997). It is the network structure (or architecture) that gives rise to the functional form of the resulting flexible-regression function.

ANN neurons are grouped in "layers." At a minimum, ANNs consist of an "input layer" and an "output layer", but may also include intermediate "hidden layers." Neurons in the hidden or output layers aggregate weighted inputs — sent from neurons in the previous layer — and transforms the aggregated value to produce a new output value. In a single-output ANN with no hidden layers, for example, the output neuron receives K inputs associated with observation i, $x_{i,k}$, weighted by a parameter, β_k, from the input-layer neurons, aggregates them to obtain a single value, "net_i", and then performs a transformation of "net_i", $\mathcal{F}(net_i)$, to produce an individual output, y_i. Here, $\mathcal{F}$ is termed an "activation function" and is commonly a sigmoid function, such as the logistic or hyperbolic tangent function (West et al., 1997). An intercept term can also be added to yield (Fausett, 1994):

$$net_i = \beta_0 + \sum_{k=1}^{K} \beta_k x_{i,k} \tag{3.26}$$

and

$$y_i = \mathcal{F}(net_i) = \mathcal{F}\left(\beta_0 + \sum_{k=1}^{K} \beta_k x_{i,k}\right). \tag{3.27}$$

Hidden layers can be added to approximate highly nonlinear functions. A researcher can think of each hidden layer as a way to reduce the dimensionality of the problem to improve the approximation capabilities of the ANN. In a single-hidden-layer network with a single output, the input-layer neurons send signals $\beta_{k,h} x_{i,k}$ to each hidden-layer neuron, where k and h denote the neurons sending and receiving the signal, respectively. Each hidden-layer neuron aggregates the input signals received to form $net_{i,h}$, which is then transformed using an activation function to obtain an output:

$$y_{i,h} = \mathcal{F}_1(net_{i,h}) \text{ for } h = 1, \ldots, H, \tag{3.28}$$

where

$$net_{i,h} = \beta_{0,h} + \sum_{k=1}^{K} \beta_{k,h} x_{i,k} \tag{3.29}$$

and $\mathcal{F}_1$ is the hidden-layer activation function. Each hidden-layer neuron then sends a signal $\delta_h y_{h,i}$ to the output layer. The output-layer neuron sums the signals plus (optionally) an intercept term, δ_0, to obtain net_i, which is then transformed using a second activation function. The resulting output is given by

$$y_i = \mathcal{F}_2(net_i), \tag{3.30}$$

where $\mathcal{F}_2$ is the output-layer transformation function and

$$net_i = \delta_0 + \sum_{h=1}^{H} \delta_h \mathcal{F}_1 \left(\beta_{0,h} + \sum_{k=1}^{K} \beta_{k,h} x_{i,k} \right). \tag{3.31}$$

While multiple hidden layers can be considered, only single-hidden-layer networks are examined in this study.

In addition to hidden layers, ANNs can also be constructed to produce multiple outputs. In this case, each of the intermediate outputs produced by the hidden-layer neurons, $y_{i,h}$, are weighted and sent to $j = 1, \ldots, J$ output neurons, where the weights are unique to each hidden neuron-output neuron pair. For this architecture, outputs are given by

$$y_{i,j} = \mathcal{F}_2 \left(net_{i,j} \right) \tag{3.32}$$

where

$$net_{i,j} = \delta_{0,j} + \sum_{h=1}^{H} \delta_{h,j} \mathcal{F}_1 \left(\beta_{0,h} + \sum_{k=1}^{K} \beta_{k,h} x_{i,k} \right). \tag{3.33}$$

If $\mathcal{F}_2$ is chosen to be a purely linear function, then $y_{i,j}$ is simply equal to equation 3.33, making $\mathcal{F}_2$ the identity function.

A primary objective of this paper is to relax the assumption that $g\left(\mathbf{x}_j; \boldsymbol{\beta}_j\right)$ (first seen in equation 3.4) is given by $\mathbf{x}_j' \boldsymbol{\beta}_j$. Instead, the ANN approach allows for a semi-nonparametric approximation of $g\left(\mathbf{x}_j; \boldsymbol{\beta}_j\right)$ by setting it equal to the right-hand side of equation 3.33. Note that in this case, we have that $g\left(\mathbf{x}_j; \boldsymbol{\beta}_j\right) = g\left(\mathbf{x}_j; \boldsymbol{\beta}_j, \boldsymbol{\delta}_j\right)$ and that $g\left(\mathbf{x}_0; \boldsymbol{\beta}_0, \boldsymbol{\delta}_0\right) = 0$ for $\left[\boldsymbol{\beta}_j' \ \boldsymbol{\delta}_j'\right] = [\mathbf{0}' \ \mathbf{0}']$, as required. Under this framework, the share of crop $j = 1, \ldots, J$ is given by

$$s_j = P_j = \frac{\exp\left(\delta_{0,j} + \sum_{h=1}^{H} \delta_{h,j} \mathcal{F}_1 \left(\beta_{0,h} + \sum_{k=1}^{K} \beta_{k,h} x_{i,k}\right)\right)}{1 + \sum_{j=1}^{J} \exp\left(\delta_{0,j} + \sum_{h=1}^{H} \delta_{h,j} \mathcal{F}_1 \left(\beta_{0,h} + \sum_{k=1}^{K} \beta_{k,h} x_{i,k}\right)\right)} \tag{3.34}$$

and thus the estimated equations from expression 3.12 become

$$\ln\left(\frac{s_j}{s_0}\right) = \delta_{0,j} + \sum_{h=1}^{H} \delta_{h,j} \mathcal{F}_1 \left(\beta_{0,h} + \sum_{k=1}^{K} \beta_{k,h} x_{i,k} \right). \tag{3.35}$$

The structure of an ANN given by equation 3.33 is represented in figure 3.1. As the figure depicts, the ANN approach, as with the SUR-HEAR, estimates the system of log-share ratios with a single model. Thus, like the SUR-HEAR, the ANN approach should capture the contemporaneous correlation between errors, as the parameterization of the ANN has common parameters across equations and the equations are jointly modeled

ANN Estimation

ANN connection weights ($\boldsymbol{\beta}, \boldsymbol{\delta}$) are often estimated using a method known as back-propagation that updates weights based on the derivative of an error function with respect to individual weights (Principe, Euliano, & Lefebvre, 2000). A common error function — and the one used in this study — is the mean square error (MSE). For this study, MSE can be represented by

$$MSE = \frac{1}{J}\frac{1}{T}\frac{1}{N}\sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\hat{y}_{i,t,j} - y_{i,t,j}\right)^2, \tag{3.36}$$

where i denotes region, t denotes the time period, j denotes crop, $\hat{y}_{i,t,j}$ is the ANN-estimated output, and $y_{i,t,j} = \ln\left(\frac{s_j}{s_0}\right)$.

A concern when estimating ANNs is over-training, or fitting the data too closely (Principe et al., 2000). To protect against this, ANNs are typically trained (estimated) using two datasets. The first set — the training set — is used to train the network (i.e., update the weight values). The second set — the validation set — is used as a form of cross-validation. Tracking the performance of a network on the validation set, in terms of MSE for example, allows for the network's generalizability (or lack of over-training) to be monitored and built into a stopping rule (Principe et al., 2000). For example, in this study, ANNs are trained using MATLAB, which uses two criteria to determine when to stop training. The first component looks at the gradient of the performance, and terminates training if it is less than $1E-5$ (Beale, Hagan, & Demuth, 2017). The second component looks at the number of validation checks. If performance (MSE) on the validation set fails to decrease for six consecutive iterations, training is terminated (Beale et al., 2017). For this study, 80% of the observations are used as the training set and the remaining 20% serve as the validation set.

A second consideration is the specification of the network architecture. When determining the number of neurons to place in a hidden layer, there is no general rule which can be employed. This decision plays an important role in network performance, as too few hidden-layer neurons will result in what Principe et al. (2000, p. 250) term "model bias" whereas too many neurons results in "model variance". The former can be thought of as an under-trained network and the latter as an over-trained network. In order to find the "best" network for this study, multiple specifications are tested where the number of hidden-layer neurons changes across specifications. For each specification, the network is estimated 1,000 times with randomized subsets of the data serving as the training and validation datasets. This type of approach was suggested by Breiman (1996), who referred to it as bootstrap-bagging or simply "bagging." The preferred network specification for elasticity calculations is chosen based on the lowest average MSE — with respect to the validation data — across the 1,000 runs.

A final consideration addressed in this study is the choice of activation functions, $\mathcal{F}_1$ and $\mathcal{F}_2$ in this case. Sigmoid functions such as the hyperbolic tangent or logistic cumulative density function (cdf) are typically used, particularly in the hidden layer due to the function approximation benefits they yield (see section 3.1). In this study, the logistic cdf is used for $\mathcal{F}_1$. The problem this creates is that the error function, e.g. MSE, becomes highly nonlinear. As a result, the optimization procedure may stop when it hits a local minimum rather than the global minimum. To reduce the risk of this occurring, an ANN can be estimated multiple times with a different set of initial parameter values for each iteration. For this study, once the best network architecture is found, as described above, it is re-estimated — again using 1,000 data partitions — but now with 10 sets of initial parameters (i.e., starting values) used for training on each partition. In other words, an additional 10,000 networks are trained using the preferred specification. Each new set of starting values results in a different MSE for the validation set. For each group of 10, the network that has the lowest validation MSE is kept and used to estimate elasticity values. This approach thus provides 1,000 elasticity estimates for each crop-year-region combination that are used to estimate means and standard errors.

3.3.3 Elasticities

In general, the acreage elasticity for crop j in region i for year t with respect to variable k is given by

$$\eta_{i,t,j,k} = \frac{\partial A_{i,t,j}}{\partial x_{i,t,j,k}} \frac{x_{i,t,j,k}}{A_{i,t,j}} \tag{3.37}$$

where $A_{i,t,j}$ is the total acres in region i allocated to crop j at time t. Using land-use shares, equation 3.37 can be rewritten as

$$\eta_{i,t,j,k} = \frac{\partial s_{i,t,j} \bar{A}_i}{\partial x_{i,t,j,k}} \frac{x_{i,t,j,k}}{s_{i,t,j} \bar{A}_i} \tag{3.38}$$

where $\bar{A}_i$ is the total potential agricultural land in the region. Then, under the assumption that $s_{i,t,t} = P_{i,t,j}$, this becomes

$$\eta_{i,t,j,k} = \frac{\partial P_{i,t,j} \bar{A}_i}{\partial x_{i,t,j,k}} \frac{x_{i,t,j,k}}{P_{i,t,j} \bar{A}_i} \tag{3.39}$$

or

$$\eta_{i,t,j,k} = \frac{\partial P_{i,t,j}}{\partial x_{i,t,j,k}} \frac{x_{i,t,j,k}}{P_{i,t,j}}, \tag{3.40}$$

where $\frac{\partial P_{i,t,j}}{\partial x_{i,t,j,k}} = ME_{i,t,j,k}$ is the marginal effect with respect to variable $x_{i,t,j,k}$ on the probability of observing crop j in a particular field in region i at time t.

With the SUR-HEAR approach, $ME_{i,t,j,k}$ is given by

$$ME_{i,t,j,k} = P_{i,t,j} \left(\beta_{j,k} - \sum_{j=1}^{J} \beta_{j,k} P_{i,t,j} \right), \tag{3.41}$$

while for the ANN approach it is given by

$$\begin{aligned} ME_{i,t,j,k} = \; & P_{i,t,j} \Big(\textstyle\sum_{h=1}^{H} \delta_{h,j} \mathcal{F}_1 (net_{i,h}) \left[1 - \mathcal{F}_1 (net_{i,h})\right] \beta_{k,h} - \\ & \textstyle\sum_{j=1}^{J} \sum_{h=1}^{H} \delta_{h,j} \mathcal{F}_1 (net_{i,h}) \left[1 - \mathcal{F}_1 (net_{i,h})\right] \beta_{k,h} P_{i,t,j} \Big), \end{aligned} \tag{3.42}$$

where $net_{i,h}$ is as given in equation 3.29. Using equations 3.39-3.42, the elasticities become

$$\eta_{i,t,j,k} = x_{i,t,j,k}\left(\beta_{j,k} - \sum_{j=1}^{J}\beta_{j,k}P_{i,t,j}\right) \tag{3.43}$$

for the SUR-HEAR model and

$$\begin{aligned}\eta_{i,t,j,k} = \quad & x_{i,t,j,k}\Big(\textstyle\sum_{h=1}^{H}\delta_{h,j}\mathcal{F}_1\left(net_{i,h}\right)\left[1-\mathcal{F}_1\left(net_{i,h}\right)\right]\beta_{k,h} - \\ & \textstyle\sum_{j=1}^{J}\sum_{h=1}^{H}\delta_{h,j}\mathcal{F}_1\left(net_{i,h}\right)\left[1-\mathcal{F}_1\left(net_{i,h}\right)\right]\beta_{k,h}P_{i,t,j}\Big)\end{aligned} \tag{3.44}$$

for the ANN model.

To find the elasticity at an aggregated level, say county or state, the same basic approach can be used. In this case, for a particular region r, the total potential agricultural land can be calculated as

$$\bar{A}_r = \sum_{i=1}^{N}\omega_{i,r}\bar{A}_i, \tag{3.45}$$

where $\omega_{i,r} \in [0,1]$ is the share of the smaller unit i which lies within the larger unit r. Similarly, the total acreage allocated to crop j in the aggregated region can be calculated as

$$A_{r,t,j} = \sum_{i=1}^{N}\omega_{i,r}s_{i,t,j}\bar{A}_i \tag{3.46}$$

or, using the assumption that $s_{i,t,j} = P_{i,t,j}$,

$$A_{r,t,j} = \sum_{i=1}^{N}\omega_{i,r}P_{i,t,j}\bar{A}_i. \tag{3.47}$$

The marginal effect on total acreage in region r in year t allocated to crop j with respect to variable k can then be calculated as

$$ME_{r,t,j,k} = \frac{\partial A_{r,t,j}}{\partial \mathbf{x}_{r,t,k}} = \sum_{i=1}^{N}\frac{\partial \omega_{i,r}P_{i,t,j}\bar{A}_i}{\partial x_{i,t,j,k}} \tag{3.48}$$

or

$$ME_{r,t,j,k} = \sum_{i=1}^{N} \omega_{i,r} \bar{A}_i \frac{\partial P_{i,t,j}}{\partial x_{i,t,j,k}}. \tag{3.49}$$

The elasticity for the aggregated region can then be calculated as

$$\eta_{r,t,j,k} = \left[\frac{\partial \bar{A}_{r,t,j}}{\partial \mathbf{x}_{r,t,k}}\right]' \frac{\mathbf{x}_{r,t,k}}{\bar{A}_{r,t,j}} = \frac{1}{\bar{A}_{r,t,j}} \sum_{i=1}^{N} \omega_{i,r} \bar{A}_i \frac{\partial P_{i,t,j}}{\partial x_{i,t,j,k}} x_{i,t,j,k}, \tag{3.50}$$

where $\mathbf{x}_{r,t,k} = [x_{1,t,j,k} \cdots x_{i,t,j,k} \cdots x_{N,t,j,k}]$. For the SUR-HEAR approach, equation 3.50 becomes

$$\eta_{r,t,j,k} = \frac{1}{\bar{A}_{r,t,j}} \sum_{i=1}^{N} \omega_{i,r} \bar{A}_i P_{i,t,j} \left(\beta_{j,k} - \sum_{j=1}^{J} \beta_{j,k} P_{i,t,j}\right) x_{i,t,j,k} \tag{3.51}$$

and for the ANN approach becomes

$$\begin{aligned} \eta_{r,t,j,k} = \; & \tfrac{1}{\bar{A}_{r,t,j}} \textstyle\sum_{i=1}^{N} \omega_{i,r} \bar{A}_i P_{i,t,j} \left(\sum_{h=1}^{H} \delta_{h,j} \mathcal{F}_1 \left(net_h\right) \left[1 - \mathcal{F}_1 \left(neth_h\right)\right] \beta_{k,h} - \right. \\ & \left. \textstyle\sum_{j=1}^{J} \sum_{h=1}^{H} \delta_{h,j} \mathcal{F}_1 \left(net_h\right) \left[1 - \mathcal{F}_1 \left(net_h\right)\right] \delta_{h,j} P_{i,t,j} \right) x_{i,t,j,k}. \end{aligned} \tag{3.52}$$

An alternative, more compact way of writing equations 3.51 and 3.52 is

$$\eta_{r,t,j,k} = \sum_{i=1}^{N} \gamma_{i,r,j} \eta_{i,t,j,k} \tag{3.53}$$

where

$$\gamma_{i,r} = \frac{\omega_{i,r} \bar{A}_i P_{i,t,j}}{\bar{A}_{r,t,j}} \tag{3.54}$$

and $\eta_{i,t,j,k}$ is the sub-region elasticity given in either equation 3.43 or 3.44. In other words, the regional elasticity can be viewed — and calculated — as a weighted average of the sub-region elasticities, where the weights are the crop-acreage shares in region r contributed by each sub-region.

Standard errors for the elasticity estimates can be obtained with either the delta method or a bootstrap approach. For this study, standard errors are estimated via bootstrapping. For the ANNs, the bagging procedure previously described serves as the method for calculating the standard errors. SUR-HEAR standard errors are approximated by performing the procedure 1,000 times using random subsets of the data where sampling is done with replacement. In order to perform the SUR-HEAR procedure, the full history of the township was needed. Thus, if a township was randomly selected, all nine years of its data were retained. The number of selected township units for each bootstrapped model was 1,873 (80%).

Based on previous acreage-response studies, it is expected that own-price acre elasticities will be inelastic. Table 3.1 provides some results from the literature. None of the elasticities in the table exceed 1.0, and many are 0.10 or less. As expected, own-price elasticities in the table tend to be positive, except for wheat from Bridges and Tenkorang (2009).

3.3.4 Model Comparisons

The SUR-HEAR and ANN approaches are compared based on model fit and elasticity results. For model fit, the two approaches are compared based on the MSE with respect to predicted acre shares — not the log-share ratios — and on the accuracy of total acreage estimates for the state. For the SUR-HEAR as well as individual ANN runs, MSE for land use j in year t is calculated as:

$$MSE_j = \frac{1}{N}\sum_{i=1}^{N}\left(s_{i,t,j} - \hat{s}_{i,j,t}\right)^2, \tag{3.55}$$

where $N = 2,342$. To get an average MSE for crop j across all years, equation 3.55 becomes:

$$MSE_j = \frac{1}{N \times T}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(s_{i,t,j} - \hat{s}_{i,j,t}\right)^2, \tag{3.56}$$

where $T = 9$ for the SUR-HEAR and $T = 8$ for the ANNs. A similar adjustment can be made to find the average MSE across crops for a given year t by replacing T with $J + 1 = 5$. An average MSE across all estimated ANNs for crop j in year t is calculated as:

$$MSE_j = \frac{1}{1000 \times N}\sum_{m=1}^{1000}\sum_{i=1}^{N}\left(s_{i,j,t,m} - \hat{s}_{i,j,t,m}\right)^2. \tag{3.57}$$

Similar adjustments can be made to equation 3.57 to obtain an average MSE across networks and years for a particular crop, or across networks and crops for a particular year. Finally, to get an average across crops and years, we can use:

$$MSE_j = \frac{1}{N \times T \times J} \sum_{j=0}^{4} \sum_{t=1}^{T} \sum_{i=1}^{N} (s_{i,t,j} - \hat{s}_{i,j,t})^2 \tag{3.58}$$

in the case of the SUR-HEAR or individual ANN and

$$MSE_j = \frac{1}{1,000 \times N \times T \times J} \sum_{i=1}^{N} (s_{i,t,j} - \hat{s}_{i,j,t})^2 \tag{3.59}$$

for an average across the 1,000 ANNs.

To make comparisons on acreage estimates, total acreages for each land use are estimated by using equation 3.47 where $\omega_{i,r} = 1$ for all i (each township unit falls completely with the state). Estimated acreages are then used to calculate the deviations from the actual acreages using

$$D_{t,j} = \frac{\hat{A}_{r,t,j}}{A_{r,t,j}} - 1. \tag{3.60}$$

Thus, $D_{t,j} > 0$ indicates the total acres in the state for crop j in year t were overestimated whereas $D_{t,j} < 0$ indicates the acres were underestimated. The values $D_{t,j}$ can then be used to calculate the mean-absolute deviation (MAD) for either a crop across years or a year across crops.

3.4 Data

Kansas land-use data is used to compare the two approaches outlined in section 3.3. Counties are a common unit-of-analysis choice in land-use studies (e.g., Lichtenberg (1989), Wu and Brorsen (1995), Lubowski et al. (2006), etc.). However, given the availability of spatially explicit data, this approach may miss an opportunity to capitalize on spatial variation in variables. Thus, this study uses a finer spatial resolution. Specifically, this study uses Public Land Survey System (PLSS) boundaries as the unit-of-analysis. The PLSS divides land into six-square-mile regions wherein a section — one square mile — is identified by section, township, and range identifiers. The 6-square mile divisions are referred to simply as townships and serve as the unit-of-analysis for this study. These regions, of which there are 2,344 in all, are depicted in figure 3.2. As can be seen in the figure, the aggregation on township and range identifiers does not create a perfectly gridded set of units. Additionally, because some townships cross state lines, they are "clipped" to the state border and thus have a much smaller area. The average township size is 35.1 mi.^{2}, or roughly 22,449 acres.

A number of explanatory variables are included in the empirical models: 14 (plus a constant) for each equation in the SUR-HEAR model and 22 for the ANN models, i.e, there are 22 input-layer neurons. The included variables are meant to capture influences from markets; climate; and soil productivity. Additionally, lagged land-use shares are included in the ANN models to account and correct for the temporal correlation. The land-use share and explanatory variable data are detailed in subsections 3.4.1 to 3.4.4 below; summary statistics are found in tables 3.2 to 3.5.

3.4.1 Land-use Shares

Land-use share values were calculated for five land-use categories: corn, sorghum, soybeans, wheat, and an "all other potential agricultural land" category. Areas were assigned to each township unit for each of the five categories based on Cropland Data Layer (CDL) raster images from the United States Department of Agriculture, National Agricultural Statistics Service (NASS) for the years 2007 to 2015. Two township units are dropped from the empirical estimations due to zero acreages for corn, sorghum, soybeans, and wheat for all years. There are thus 21,078 observations per crop for the SUR-HEAR model and 18,736 for ANN models (due to a lag). An example of CDL imagery is provided in figure 3.3 and average shares (across township units) by year in table 3.2. The "other" category is comprised of CDL classification codes which were deemed to be "agricultural" or "potentially agricultural". A notable exception is the grassland category. The grassland areas were left out largely due to the fact that the CDL classification of "Grassland/Pasture" relies heavily on the United States Geological Service's (USGS) National Land Cover Database (NLCD). The NLCD has been released for the years 1992, 2001, 2006, and 2011; providing a snapshot every five years of major trends. Additionally, NASS states that "pasture and grass-related land cover categories have traditionally had very low classification accuracy in the CDL" (CDL FAQs). Table 3.3 provides a description of the CDL classification codes; their approximate acreage by year; and how they were classified for this study.

3.4.2 Economic Variables

Output prices for corn, sorghum, soybeans, and wheat represent the expected price to be received at the time of harvest. The general form of the expected prices follows that from Hendricks et al. (2014) and is given by

$$E\left(p_{t,j}\right) = FP_{t,j} + E\left(B_{t,j}\right). \tag{3.61}$$

The first component, $FP_{t,j}$, is a futures price for crop j at time t. This component is calculated as the average futures price for the crop across those months — in year t — during which the crop is typically planted. The second component, $E\left(B_{t,j}\right)$, represents an expectation of what the harvest basis will be for crop j in year t. Expected basis is set equal to the basis from the previous harvest for each crop.

Spot prices used in the expected basis calculations represent data from 961 elevators across Kansas, the locations of which are depicted in figure 3.4. Expected prices were calculated first at the elevator unit, and each township unit was then assigned the expected price associated with the nearest elevator. Because data were not available for each crop at every elevator for every year, the expected price for a township unit may be taken from different elevators across years or potentially within the same year when looking across crops.

Due to data availability, the second set of economic variables — input prices — is the only set with no spatial variation. Two input prices are considered in this study: diesel and labor. Price indices for both inputs were obtained from the NASS QuickStats on-line database (USDA-NASS). For each input, a single index value is applied to all township units in a given year. See table 3.4 for summary data on price variables.

3.4.3 Soils Variables

Field productivity, or potential productivity, is an important component in determining land-use patterns. Lichtenberg (1989) for example notes that water-sensitive crops such as corn and soybeans tend to be grown on very high qualities of land. Physical soil characteristics, in turn, are an important component in determining the productive potential of a field. The composition of a soil in terms of sand, silt, and clay has been shown to be an important factor in the level of organic carbon in the top level of soil (Burke et al., 1989). Two of these variables — the percents of clay and silt — are used in the SUR-HEAR and ANN models, while the percent of sand is dropped due to linear dependence ($\%\text{Sand} = 1 - \%\text{Clay} - \%\text{Silt}$). Two additional soil variables were included to capture the erodibility of the land. The first is called a t-factor and gives the maximum amount of soil erosion — in tons per acre — a soil can experience before crop productivity is significantly affected (USDA-NRCS, 2017). The second, *WEI*, is an index that gives a tons-per-acre-per-year value that is used in determining wind erosion (USDA-NRCS, 2014). *WEI* is assigned to a soil based on its membership in one of nine wind erodibility groups, where a higher group number is less susceptible to wind erosion (Oregon Department of Environmental Quality, 2005). The index is based on soil texture and the effect of dry soil aggregates on potential erosion rates and has a maximum value of 310 tons/acre/year for a wide, barren field (Oregon Department of Environmental Quality, 2005).

3.4.4 Weather Variables

Included weather variables are meant to capture two factors: delayed planting effects and expectations about growing season weather. To capture the first, variables on the total precipitation, average daily minimum temperature, and average daily maximum temperature over the planting season are included. For corn, soybeans and sorghum, these variables are the same while a separate set of variables is used for wheat. To proxy for producer weather expectations, a three-year average of growing season precipitation, AVG_{CSS} is included. The growing season for corn, sorghum, and soybeans was defined as April through August. The variable AVG_W provides the wheat counterpart and is based on the months of November in year $t-1$ to June in year t. These variables provide the three-year average of total precipitation over the defined growing seasons, in millimeters (mm). The variables $PREC_{CSS}$ and $PREC_W$ give the total planting window precipitation (mm), defined as May-June and September-October, respectively. $TMAX_{CSS}$, $TMAX_W$, $TMIN_{CSS}$, and $TMAX_W$ are the planting season temperature variables. These variables give the average daily maximum and minimum temperatures (C°). Data for these variables were obtained from the PRISM Climate Group (2013). The daily weather data from PRISM is provided at a four-square-kilometer scale that is interpolated based on weather stations located throughout the country. Variables were calculated as a weighted average of the PRISM grid cells falling within a township unit where the weights were equal to the percent of a township covered by a particular grid cell.

3.5 Results

3.5.1 Specification Tests

To motivate the SUR-HEAR procedure, the data were tested for the presence of autocorrelation, heteroskedasticity, and contemporaneous correlation. Autocorrelation was tested using the Lagrange-multiplier test from W. H. Greene (2012, p. 962). The null hypothesis of no autocorrelation was rejected at the 5% level for the corn equation and at the 7.5% level for the remaining equations. A null hypothesis of no heteroskedasticity was rejected at the 1% level for each crop based on the Lagrange-multiplier test from W. H. Greene (2012, p. 316). A null hypothesis of no contemporaneous correlation was also rejected at the 1% level based on the Lagrange-multiplier test from W. H. Greene (2012, p. 338). Test statistics are presented in table 3.6.

To motivate the use of the ANNs, the Ramsey (1969) RESET test was used as described in W. H. Greene (2012, p. 177). The RESET test is a two-step procedure that can be used to assess the linearity assumption of the standard approach wherein $g\left(\mathbf{x}_j;\boldsymbol{\beta}_j\right)=\mathbf{x}_j'\boldsymbol{\beta}_j$. In the first step, the SUR-HEAR model is estimated and used to obtain fitted values via $\hat{y}_j=\mathbf{x}_j'\hat{\boldsymbol{\beta}}_j$. In the second step, the SUR model is estimated again — using the already transformed data — with two additional terms, $\hat{y}_j^2$ and $\hat{y}_j^3$ included as regressors. The null hypothesis of the null model is then accepted or rejected by simply looking at the significance of the coefficients on these variables. As seen in table 3.7, the linearity assumption is strongly rejected for each crop equation. While the RESET test is insightful regarding the linearity assumption, it is nonconstructive: it provides no guidance on what may be the correct model (W. H. Greene, 2012). Due to their approximation capabilities, ANNs lessen the need for researchers to isolate the correct specification.

3.5.2 Best ANN Specification

A total of 45 ANN specifications were tested. The only change across specifications was the number of neurons in the hidden layer, which ranged from 1 to 45. The network specification chosen for further analysis, such as elasticities and fit comparisons, was based on the best-average performance with respect to the validation data. Using this criteria, the preferred specification included 41 neurons in the hidden layer. The average MSE across the 1,000 data partitions for 41 hidden-layer neurons was 0.6736[2]. ANN performance across the number of hidden-layer neurons is presented in figure 3.5.

3.5.3 Model Fit

Model fit favors the ANN approach for the measures examined here: MSE — with respect to predicted land-use shares — and predicted aggregate acres for the state. MSE results for the actual and predicted shares are presented in table 3.8 and figure 3.6 for the SUR-HEAR model and table 3.9 and figure 3.7 for the ANN. The values in these tables and figures were

[2]This MSE is with respect to the log-share values and does not include performance on the omitted "other" category.

calculated using the methods outlined in equations 3.55 to 3.59. Across all crops and years, the SUR-HEAR MSE was 0.021. Averaged over the 1,000 ANN runs, the MSE across all crops and years was 0.005. The lowest MSE for the SUR-HEAR was 0.005, seen in 2012 for sorghum; the largest was 0.051 for the "other" category in 2013. With the ANNs, the smallest MSE was 0.002, seen in multiple years for sorghum, while the largest was 0.01 for the 2013 "other" estimates. For both approaches, the smallest MSEs averaged across years were for sorghum (0.006 for SUR-HEAR; 0.002 for ANNs) while the largest were for wheat (0.032 for SUR-HEAR; 0.006 for ANNs).

Though the SUR-HEAR MSE results are larger than those from the ANNs, a modeler may still assume they are usable. However, further investigation shows this to be misleading. When looking at the prediction errors with respect to the estimated individual township shares, $s_{i,t,j} - \hat{s}_{i,t,j}$, a clear pattern emerges. As seen in figures 3.8 to 3.12, the linear-in-parameters and explanatory variables SUR-HEAR approach tended to over-estimate shares when actual shares are low and under-estimate shares when actual shares are high. For example, when the actual acre shares belonging to corn were in the range $(0, 0.1]$, the median residual produced by the SUR-HEAR approach was approximately -0.1, which is given by the red line inside of the blue box. The bottom and top of the blue boxes indicate the 25th and 75th percentiles, respectively.

These results can be explained by looking at patterns in the predicted values, as shown in figures 3.13 to 3.17. The figures show a lack of variation in predicted values in some of the SUR-HEAR estimates. This is particularly true for estimated corn shares, where the medians are clustered around 15% regardless of the actual shares. Wheat — the dominant category — shows a similar pattern. Predicted wheat shares rise along with actual shares until actual shares are roughly 30%, but then plateau at a predicted share of roughly 40%. The best prediction trends appear to be seen with soybeans, where predicted shares (more or less) rise as actual shares rise.

Similar analysis is presented for the best-fitting ANN[3] (BANN) in figures figures 3.18

[3] "Best-fitting" here is with respect to the 1,000 additional estimates of an ANN with 41 hidden-layer neurons trained with 10 different sets of starting values.

to 3.27. The BANN was chosen from the 1,000 estimated ANNs based on the lowest MSE across all observations. A few reasons are offered for the use of the BANN in these figures. First, to aid with interpretation: the number of outliers[4], indicated by red crosses, will likely increase if all 1,000 network results are used, which could be misleading. Second, in practice a researcher is likely to select one network for the purposes of making predictions. Finally, given the low performance variation for this network specification (see figure 3.5), the remaining networks were likely to generate similar results. Generally, the BANN produced desirable results: residuals which tend to be centered around zero and predicted shares that rise along with actual shares. One exception is sorghum, which shows patterns similar to those seen with the HEAR-SUR approach.

Effects of prediction performance at the individual township level subsequently impact predictions of state-level acreages. This can be seen in figure 3.28, table 3.10, and table 3.11, which show the deviations of predicted acreages from actual acreages for the SUR-HEAR, the BANN, and the worst ANN (WANN), which is the ANN that had the largest MSE combined across land uses. For some years and crops, the SUR-HEAR predicted total acreages well, falling within a few percent of the observed values. However, it was also prone to large deviations, such as a 0.35 (35%) overestimate of soybean acres in 2013 or a 32% underestimation of wheat acres in 2010. Across all years, SUR-HEAR accuracy was best for wheat estimates, for which the mean absolute deviation (MAD) was about 8%; and worst for the "other" category where the MAD was about 22%. Generally, the BANN produced accurate acreage estimates, with the largest deviation being an 8% underestimate of corn in 2012. The BANN also performed best with respect to wheat, with a MAD of about 1%, and generally was poorest for corn, with a MAD of roughly 4%. Both approaches tended to over-estimate the "other" share and under-estimate the corn share, with the exception of 2010 corn for both approaches. Generally, though the MADs were smaller than the SUR-HEAR counterparts, the ANN approach was more prone to consistently over- or under-estimate a particular crop. Besides those already mentioned, sorghum acres tended to be under-

[4]These figures were created with MATLAB using default settings. Under these settings, outliers are defined to be points greater than $q_3 + 1.5 \times (q_3 - q_1)$ or less than $q_1 - 1.5 \times (q_3 - q_1)$ where q_1 and q_3 are the 25th and 75th percentiles of the data, respectively.

estimated while soybean and wheat acres tended to be over-estimated by the ANNs. For the most part, the WANN and BANN produced similar deviations from actual acres in terms of both magnitude and sign. The apparent robustness of predictions across the ANNs may be a strength of the cross-validation techniques employed during estimation. Additionally, prediction results suggest that the linear-in-parameters and variables specification index function used in the SUR-HEAR model may be a gross misspecification.

3.5.4 Acreage Response

Annual acreage-response elasticities were calculated using the methods in section 3.3.3. Standard errors for the ANN results were obtained via the "bagging" approach described in section 3.3.2. Standard errors for the SUR-HEAR elasticities were estimated by bootstrapping the estimates across 1,000 runs. See section 3.3.4 for details. Table 3.12 lists the bootstrapped (bagged) elasticities, averaged across years and township units. Elasticities for individual years are found in tables 3.12 to 3.19. Patterns in the individual estimates follow those for the annual-averages, so discussion below focuses on the latter.

Estimated elasticities exhibited substantial differences across the two approaches. In general, the SUR-HEAR elasticities were much larger, in terms of absolute magnitudes, and were statistically significant far more often. Of the 40 elasticities in table 3.12 associated with the SUR-HEAR model, only one was not significant at either 1, 5, or 10% level. In contrast, for the 72 associated with the ANNs, only 26 were significant, of which 12 were associated with the lagged-dependent variables. For these variables, increasing the share of acres in $t-1$ generally had a positive affect on the estimated acres for period t, with sorghum as the only exception.

Across both approaches, own-price elasticities mostly held to economic theory: increased expected prices led to increased acres. The one exception was $PWHEAT$ in the SUR-HEAR model, where the estimated elasticity was -0.15. However, for the ANN approach, none of the own-price elasticities were statistically significant (up to the 10% level) except for $PCORN$, estimated to be 0.06. It should be noted that these represent short-run elasticities;

the long-run effects may change. Of the positive own-price elasticities with the SUR-HEAR approach, the smallest was for *PCORN* at 0.66, which was still substantially larger than the ANN estimate. Sorghum and soybean own-price elasticities were much larger, at 2.15 and 6.68 respectively. With the ANN approach, these two elasticities were 0.04 and 0.03, respectively. Though they are in the low range of own-price elasticities, the results from the ANNs appear to be more in line with those from past studies (see table 3.1). All cross-price elasticities from the SUR-HEAR were significant at the 1% level except for the effect of *PWHEAT* prices on sorghum, which was significant at the 5% level. Only four (of 12) of the cross-price elasticities were significant with the ANN approach: *PSORGHUM* and *PWHEAT* on corn acres, *PCORN* on sorghum, and *PWHEAT* on soybeans. Elasticities associated with *PDIESEL* were negative and significant for sorghum and soybeans and positive and significant for wheat with the SUR-HEAR approach. *PLABOR* had a negative and significant impact on corn, soybeans and wheat and a significant positive impact on sorghum in the SUR-HEAR model. All *PDIESEL* and *PLABOR* elasticities were insignificant with the ANN approach.

All weather variables were significant at the 1% level for the SUR-HEAR model. Average growing-season precipitation (AVG_{CSS} and AVG_W) had a positive impact on corn and soybean acres and negative impacts on sorghum and wheat acres. Similarly, with the ANN approach, AVG_{CSS} had a negative and significant effect on sorghum acres and a positive and significant impact on soybean acres. The ANN approach also estimated a positive and significant elasticity for wheat with respect to AVG_{CSS}, and for corn with respect to AVG_W. The impact of AVG_W on corn may be capturing expectations regarding precipitation during the early stages of corn growth. Total planting season precipitation ($PREC_{CSS}$, $PREC_W$) was estimated to have a significant positive impact on sorghum, soybean, and wheat acres and a significant negative impact on corn acres with the SUR-HEAR model. The impacts on corn, sorghum, and soybean acres likely reflects shifts from corn to one of the other two crops during wet years as corn is generally planted earlier. $PREC_{CSS}$ with the ANN approach had a positive and significant impact on corn, sorghum, and wheat. The significant effect of $PREC_{CSS}$ on wheat, though inconsequential in terms of magnitude, is hard to rationalize as this is weather which has not yet occurred when wheat is planted. However, this variable is likely correlated with the same variable from past years, and may thus be capturing expectations about growing-season precipitation for wheat. All temperature variables were significant at the 1% level for the SUR-HEAR approach except for the effect of $TMAX_{CSS}$ on corn, which was significant at the 5% level. Soybean elasticity associated with $TMAX_{CSS}$ is worth mentioning, as it was the largest of any in the table at -7.28. If there is a correlation between high temperatures and low precipitation, this may capture prevented plantings as the result of drought. Only two significant temperature-based elasticities were found with the ANN approach: soybean acres are negatively affected by $TMAX_W$ and positively by $TMIN_{CSS}$.

Elasticity differences are also seen with respect to spatial distributions. Own-price elasticity maps, based on the average-annual elasticities at the township level, are provided in figures 3.29 to 3.36. The ANN maps use township-level acreage elasticities averaged across all networks and years; the SUR-HEAR maps are based on the full model estimates averaged across years. In general, the SUR-HEAR produces distributions which follow an east-west gradient. An exception is the SUR-HEAR map for sorghum, for which there is little in the way of a discernible pattern. One notable difference between the two approaches is seen in the corn maps. For corn, the SUR-HEAR approach indicates that negative responses to $PCORN$ are seen in the western half of the state. The ANN approach indicates predominantly positive responses in this region. Conversely, the ANN approach indicates some negative elasticities in northeast Kansas, whereas this is a region that contains some of the stronger, positive elasticities as estimated by the SUR-HEAR model. Another reversal of sorts is seen with the wheat elasticities. For these, the SUR-HEAR model indicates mostly negative responses to $PWHEAT$ in the eastern half of the state. With the ANN approach, eastern Kansas sees some of the larger positive elasticities, while negative responses are generally in the western half, though they are somewhat scattered across the state. No negative own-price soybean elasticities are found with the SUR-HEAR model. SUR-HEAR soybean elasticities do again follow a gradient though, with elasticities increasing moving from east to west. With the ANNs, negative own-price soybean elasticities are indicated for the southwest and southeast regions of the state. Larger positive elasticities for soybean acres are seen in west/southwest Kansas for the SUR-HEAR and in central/south-central Kansas for the ANNs. Given the misspecification of the SUR-HEAR index function and the predictive performance results, one may conclude that a misspecified model will result in misleading inferences regarding spatial impacts.

3.6 Conclusions

Agricultural land-use patterns are important at various scales, such as the impact of global shocks on local livelihoods or of small-scale decisions on local or regional environments. Thus, governmental policies are often created that focus on farmer incomes, trade distortions, environmental concerns, etc. Land-use trends are also complex: they are influenced by policies, local markets, weather patterns, etc. This intersection of importance and complexity is cause for concern as to whether or not an empirical model is correctly specified. It is also reason to believe some traditional approaches, when based on a simple linear model, may be misspecified. This paper uses an empirical application based on Kansas crop-share data to present artificial neural networks (ANNs) as a viable alternative to a more traditional linear-logit specification — the heteroskedastic and time-wise autoregressive seemingly unrelated regression model (SUR-HEAR) — for estimating regional crop shares. The key point of departure between the ANN approach and the SUR-HEAR approach is the probability index function upon which both are based. A linear-in-parameters and in explanatory variables index was used with the SUR-HEAR approach, while the ANN uses a non-linear flexible functional form approximation to the true underlying index function. Ramsey (1969) specification tests indicate the SUR-HEAR index function is misspecified.

Empirical results indicated significant differences between the two approaches. The ANN approach shows a superior ability to estimate observed shares (and thus actual acreages) compared with the SUR-HEAR approach. The ANN approach also produced elasticity estimates which were, for the most part, statistically insignificant and of a much smaller magnitude. This is not to suggest that the ANN-elasticity estimates are incorrect; based on the predictive capabilities one may reasonably assume the reverse is true. While strong prediction does not always lead to accurate inferencing, e.g., when a model is over-fit to the data and thus to noise in the data, the ANN estimation procedures offer some protection against this with stopping criteria that make use of cross-validation. Additionally, ANN acreage-response elasticities were inelastic with respect to all own or cross prices, which is more or less in line with previous findings as shown in table 3.1, whereas elastic responses were often produced by the SUR-HEAR model. Potentially, the inclusion of temporal lags in the ANN specification drove this result, which would suggest that rotational considerations are more influential than prices in the short-run. Differences in the spatial distribution of own-price elasticities was also observed across the two approaches. Given the complex nature of land-use decisions, plausible explanations for either set of patterns can likely be found, with the possible exception of the SUR-HEAR sorghum results. The geographic distribution of sorghum elasticities from the SUR-HEAR model did not appear to exhibit any strong spatial pattern.

Though the ANN approach appears to offer some advantages, at least in terms of prediction, it should be weighed against researcher burden. Because the ANN approach amounts to the optimization of a highly nonlinear objective function, estimation is not as straightforward as with the SUR-HEAR model. Estimation of elasticities and acreages using ANNs used a search over 45 potential network architectures — which in another application could be more — using 1,000 random data partitions. This took a substantial amount of time given the size of the data set. Once the best network was found, a bootstrapping procedure known as "bagging" was used to obtain standard errors for elasticities. Though bootstrapping was used for SUR-HEAR standard errors as well, it took considerably less time. Another consideration to keep in mind is that a rather naively specified index function was used for the SUR-HEAR model; and performance could likely be improved through the inclusion of nonlinear terms, interactions, etc. However, in this case the traditional approach will place an additional burden on the researcher, who must now test and compare multiple specifications. Results from this study suggest that a researcher could potentially ease the ANN burdens by foregoing network specifications with a small number of hidden-layer neurons;

particularly for multiple-output specifications. In this case, the MSE with respect to the validation data set saw quick drops until reaching four hidden-layer neurons — the number of output neurons.

The true underlying functional form for land-use shares will likely never be known, placing the onus on the researcher to decide whether the results here merit the use of ANNs over a more traditional and simpler model. However, *because* the true underlying function will likely never be known, it seems appropriate to consider the ANN approach as it reduces misspecification risk. The predictive capabilities lend some support to this. Additionally, given the time which is spent collecting and preparing data, some extra time estimating should not be reason to avoid a particular approach. Since it is ultimately the duty of the researcher to enable others to develop informed opinions or make justified decisions, it may be worth assigning the task to a method which may be a little less familiar or even more difficult, such as artificial neural networks.

Figures

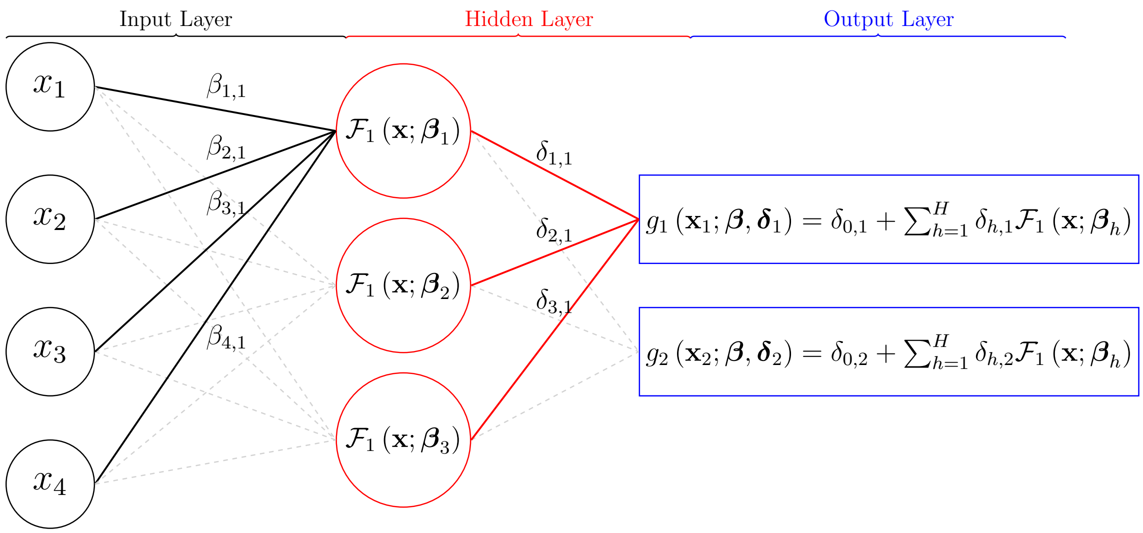

Figure 3.1: Log-acre-share ANN architecture

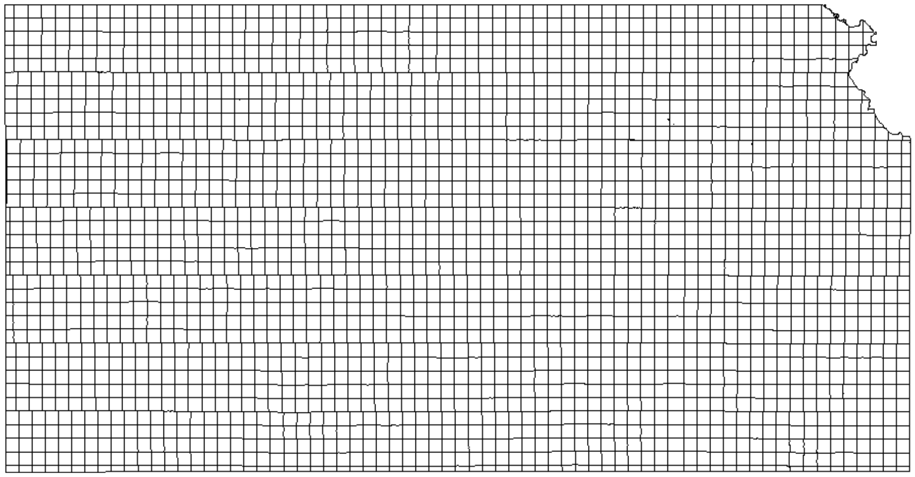

Figure 3.2: Kansas township divisions

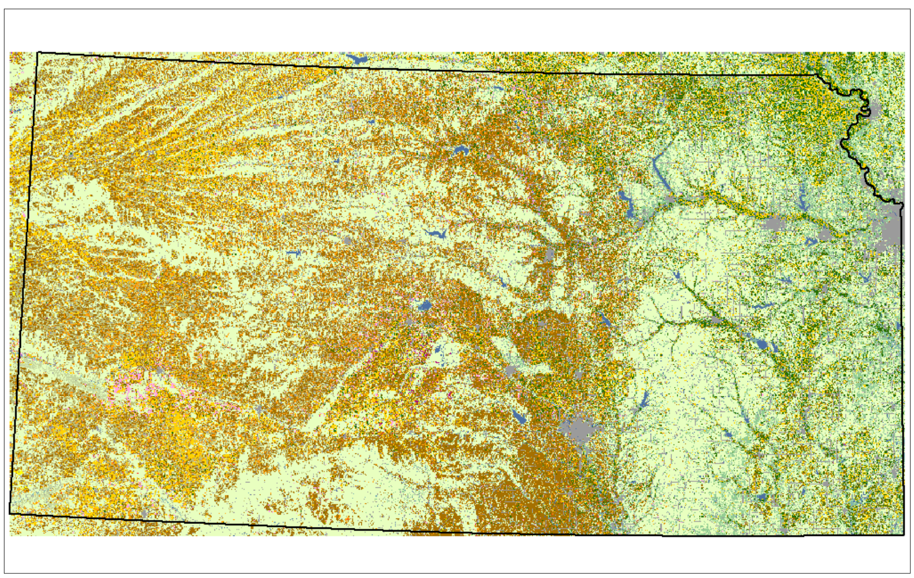

Figure 3.3: Cropland Data Layer example (2008)

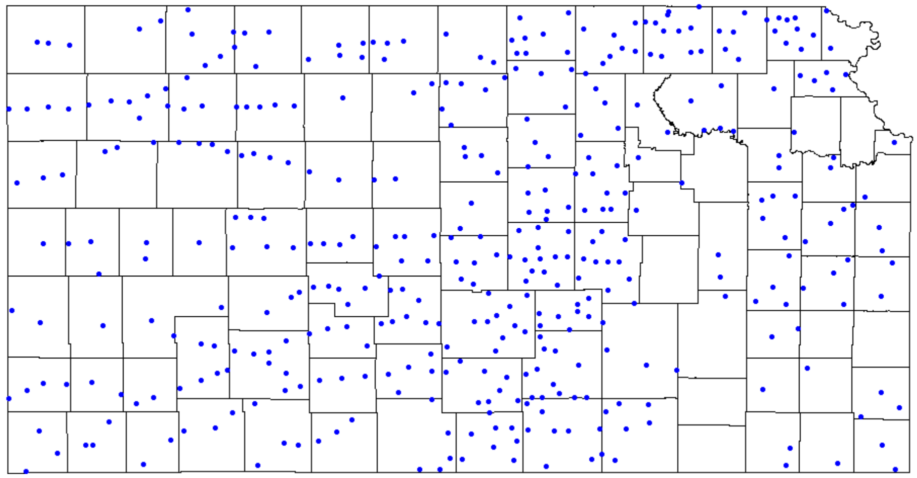

Figure 3.4: Elevator Locations

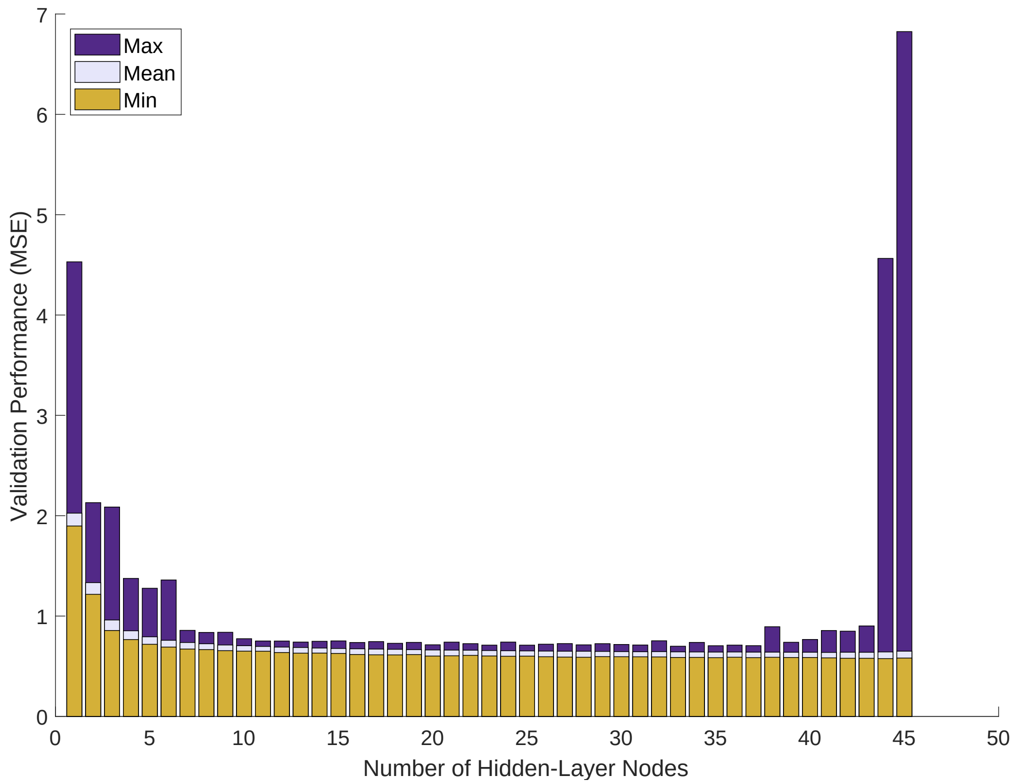

Figure 3.5: ANN performance across specifications

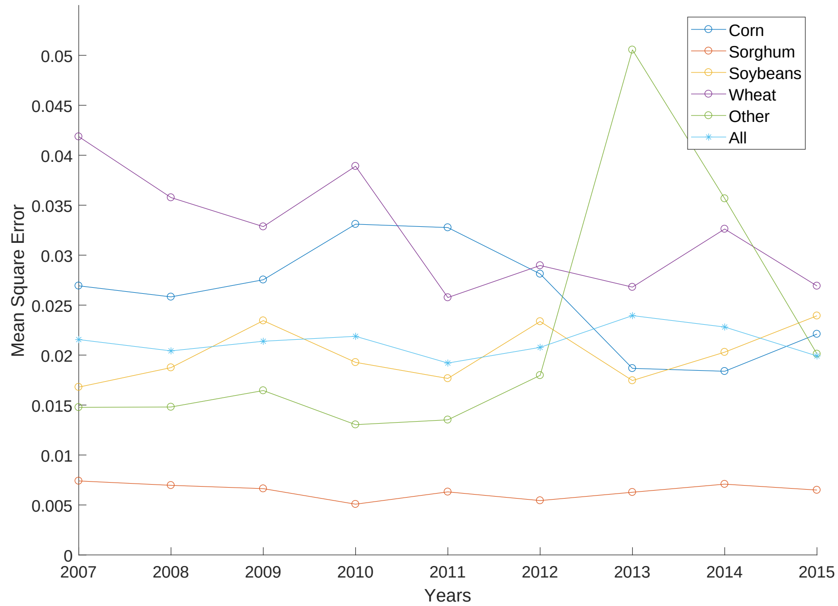

Figure 3.6: Mean square error, SUR-HEAR

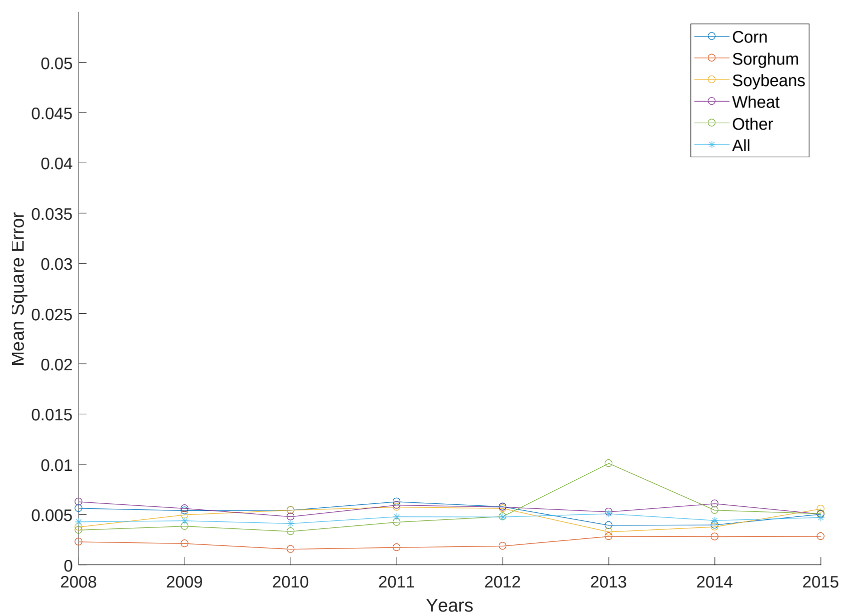

Figure 3.7: Average-Mean square error, ANNs

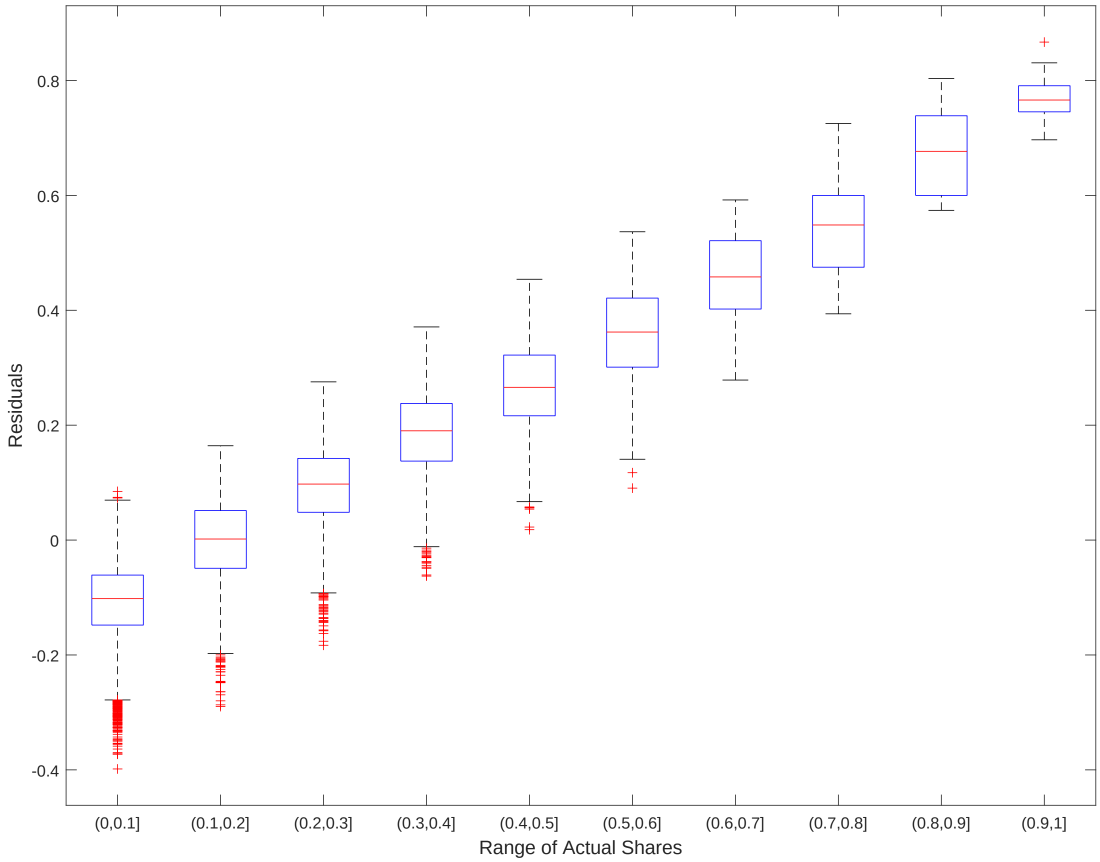

Figure 3.8: Corn residuals box-plot, SUR-HEAR

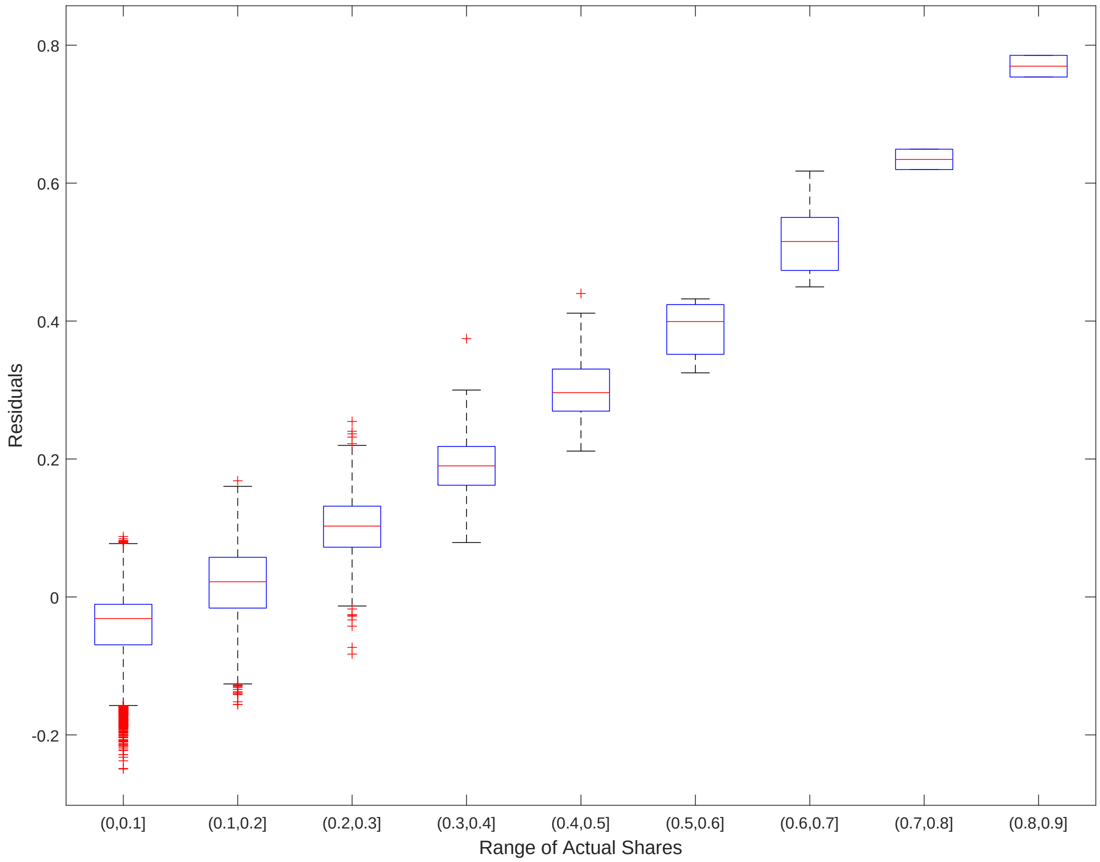

Figure 3.9: Sorghum residuals box-plot, SUR-HEAR

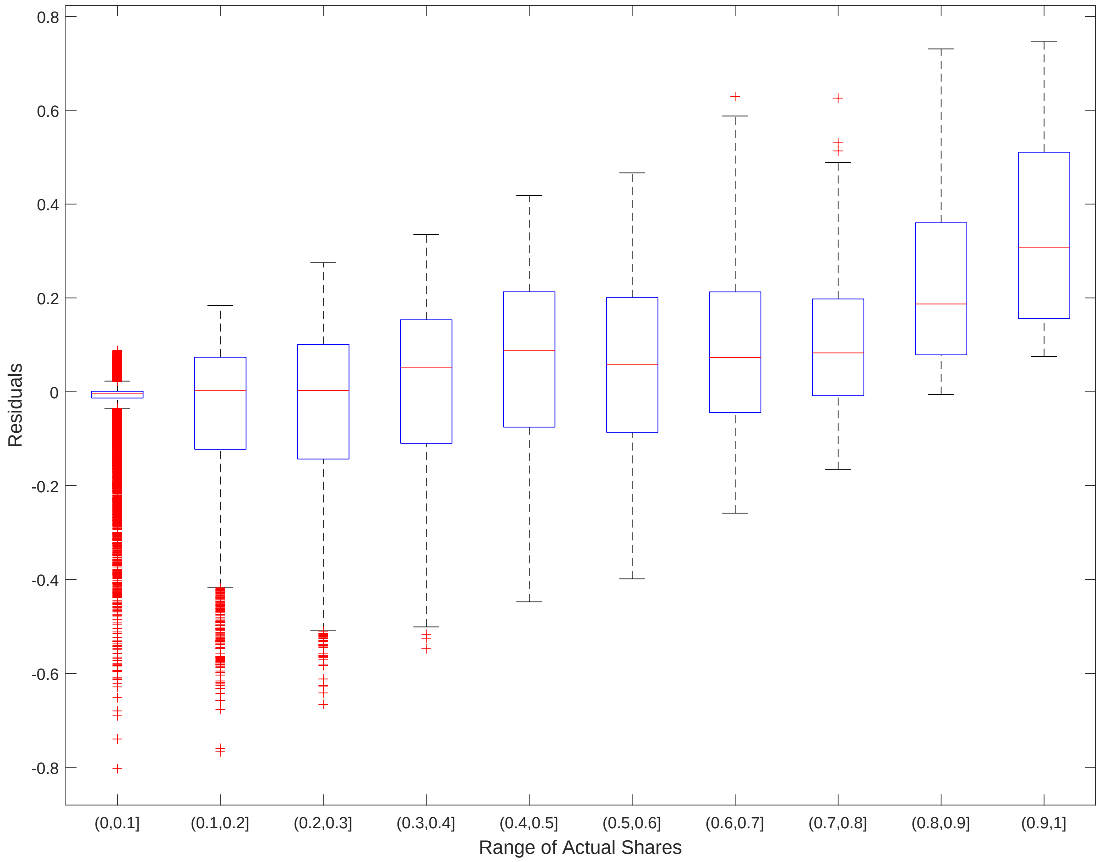

Figure 3.10: Soybean residuals box-plot, SUR-HEAR

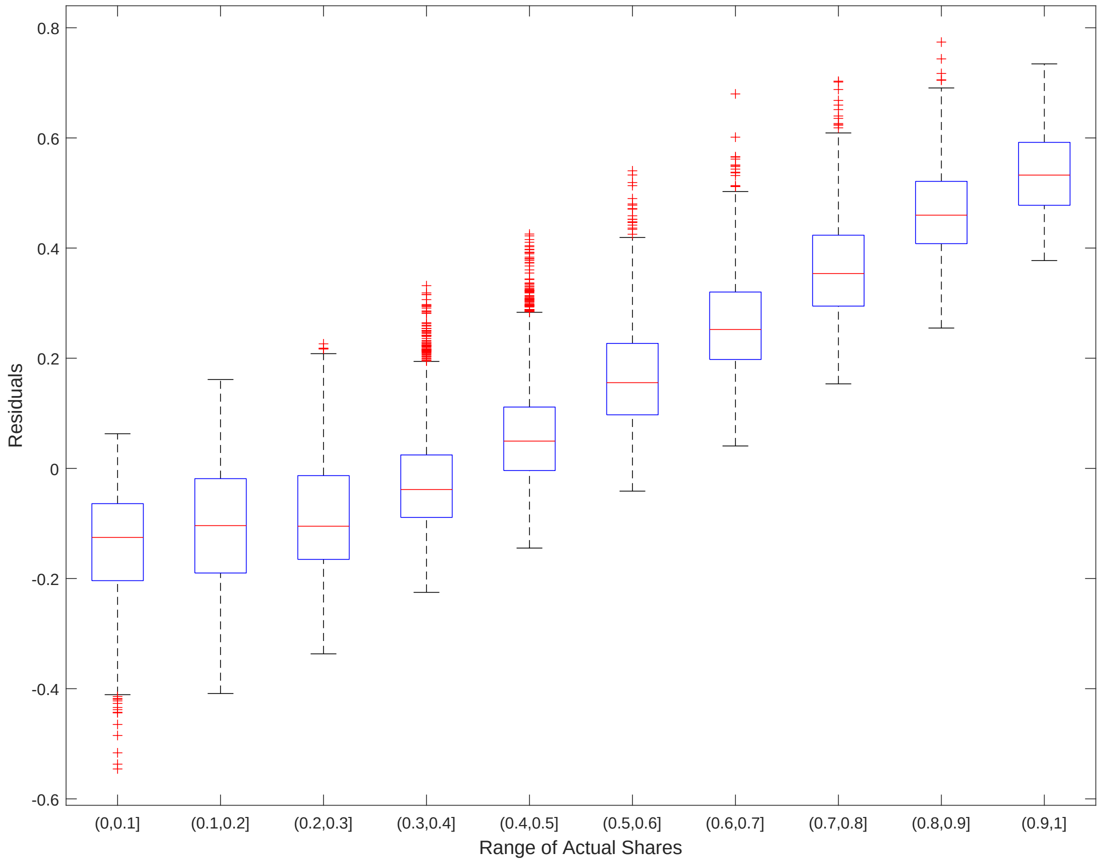

Figure 3.11: Wheat residuals box-plot, SUR-HEAR

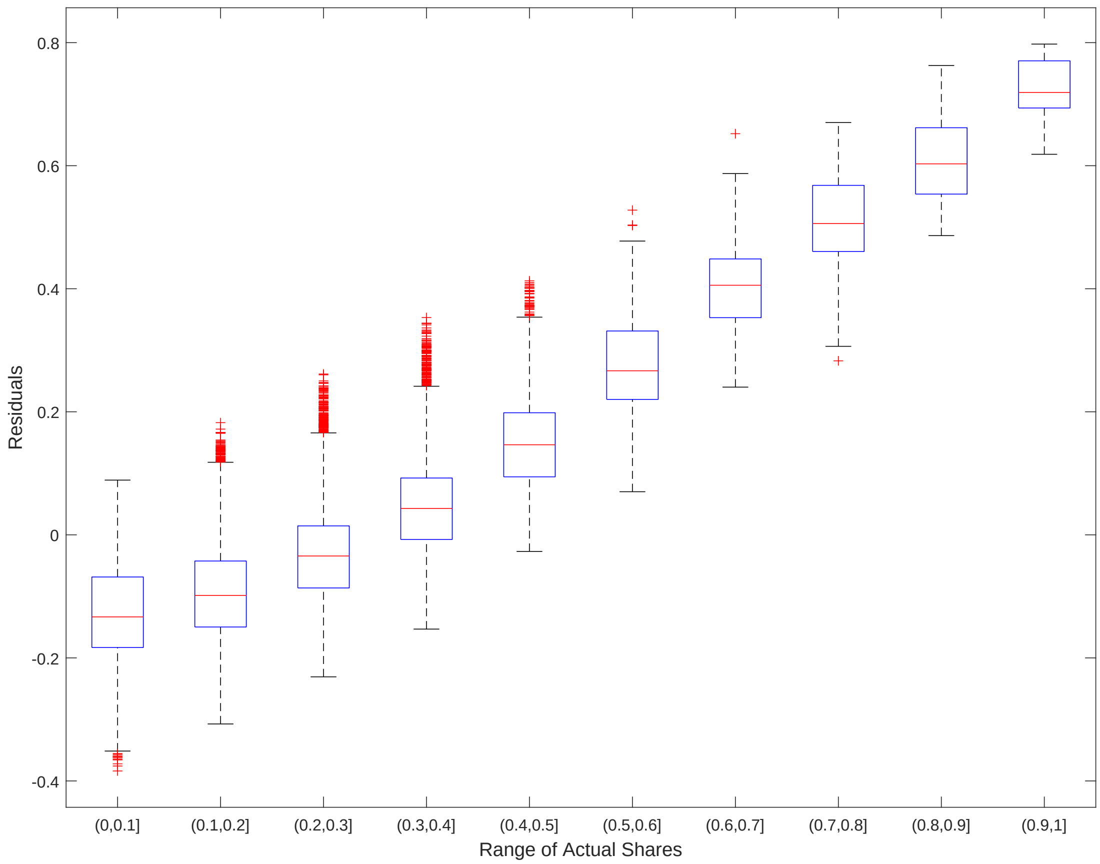

Figure 3.12: Other residuals box-plot, SUR-HEAR

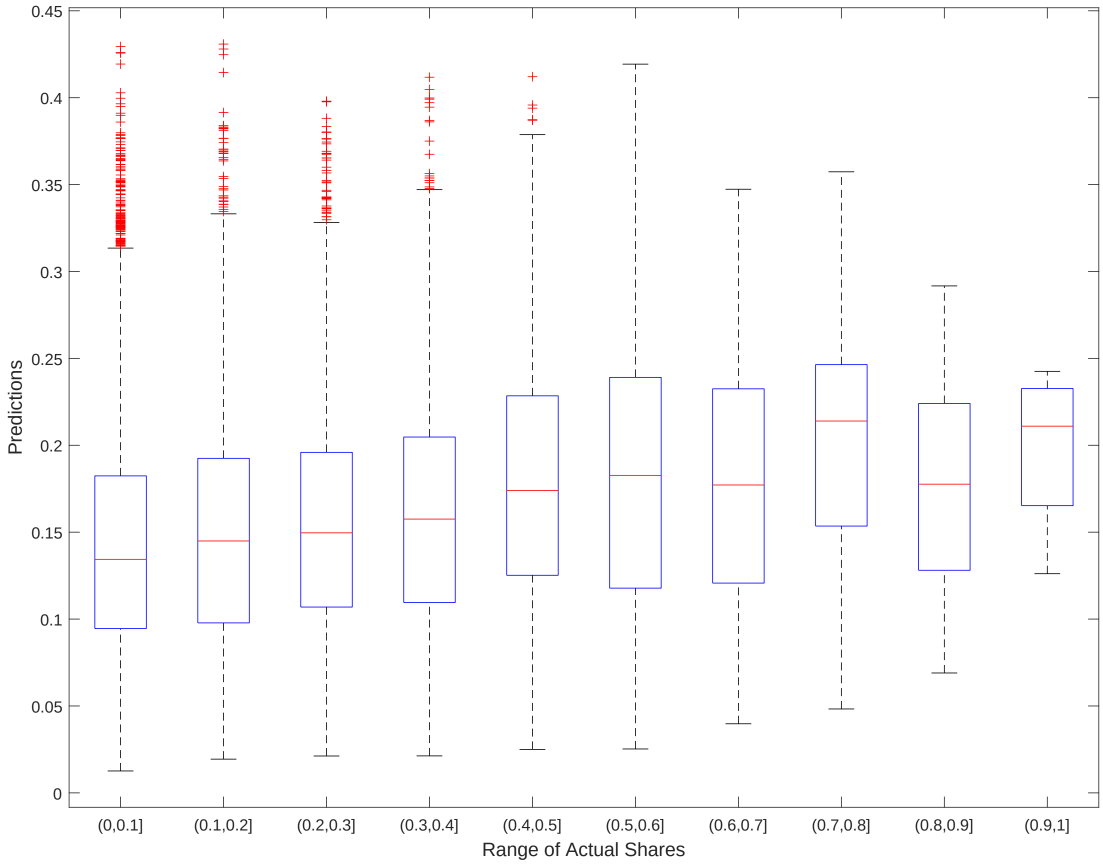

Figure 3.13: Corn predictions box-plot, SUR-HEAR

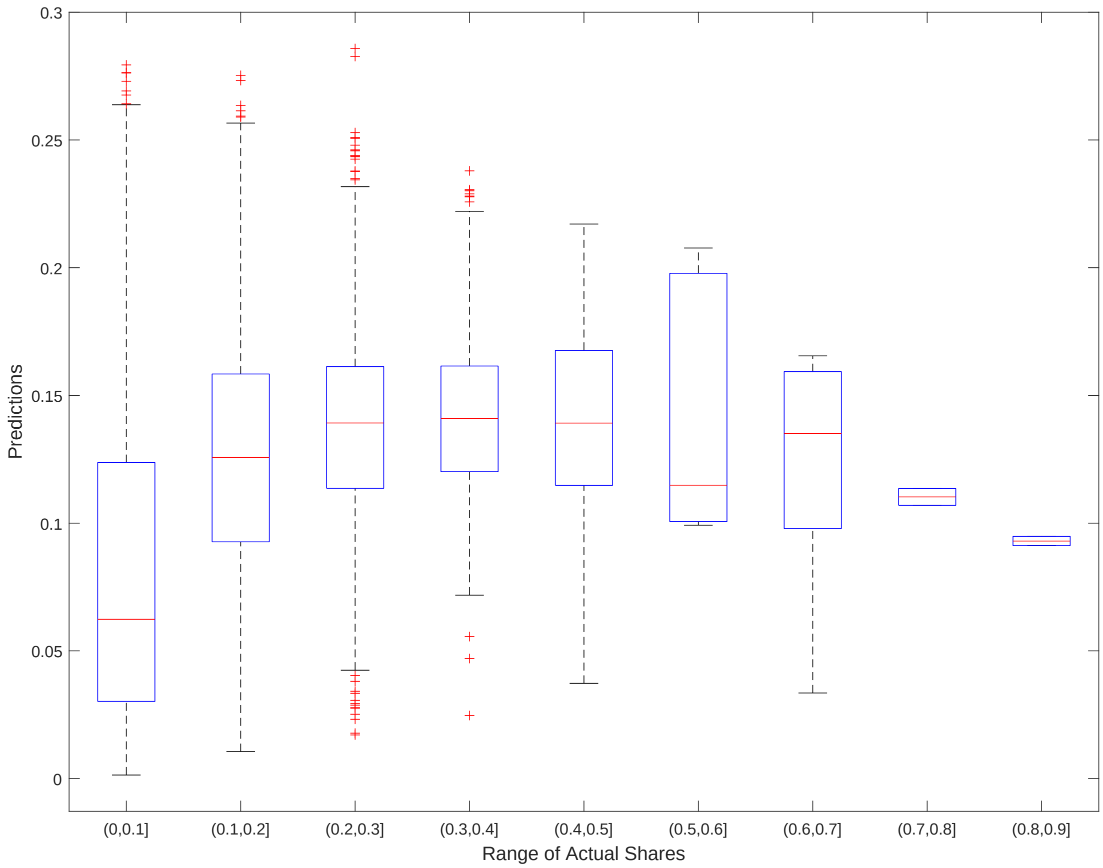

Figure 3.14: Sorghum predictions box-plot, SUR-HEAR

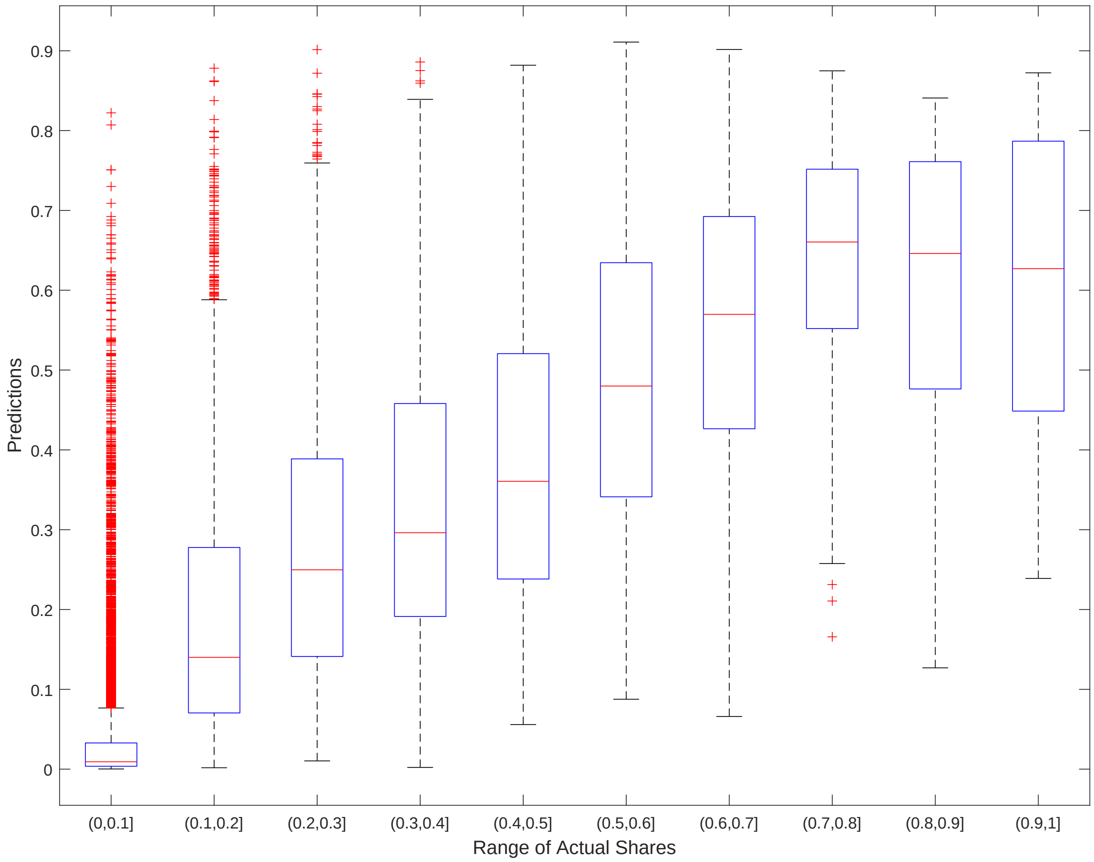

Figure 3.15: Soybeans predictions box-plot, SUR-HEAR

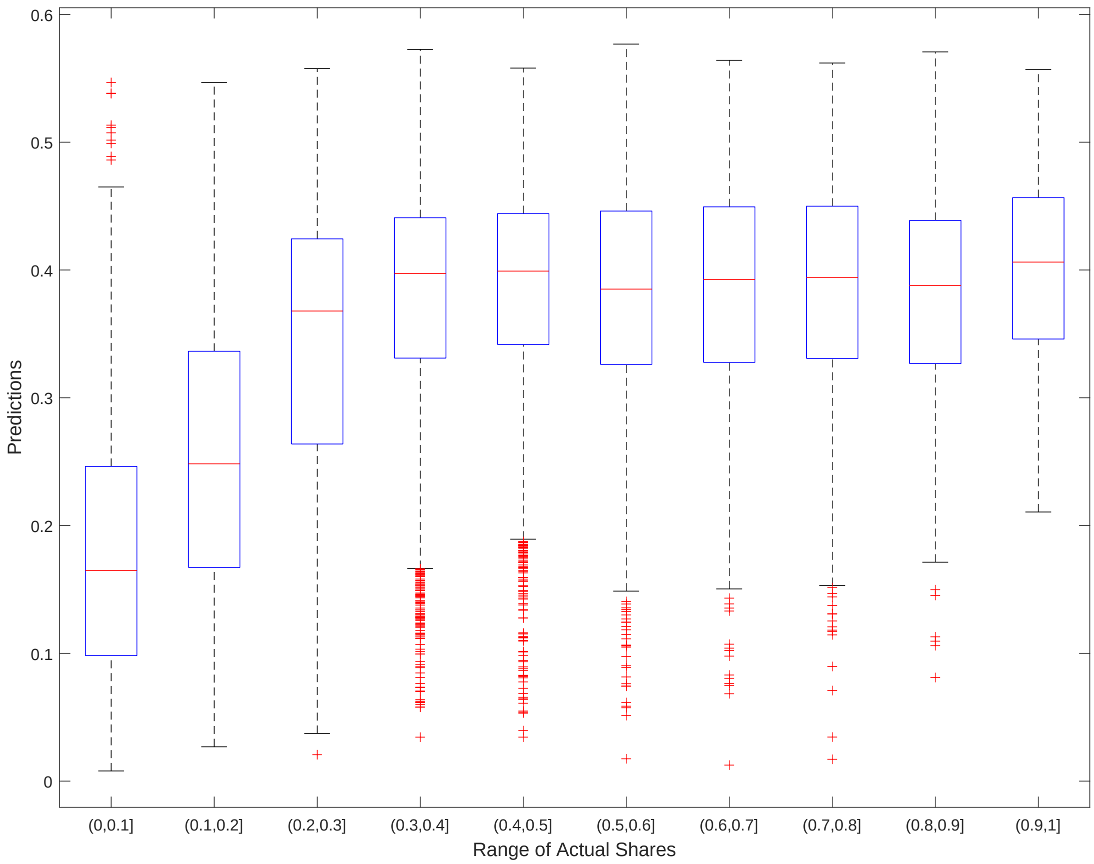

Figure 3.16: Wheat predictions box-plot, SUR-HEAR

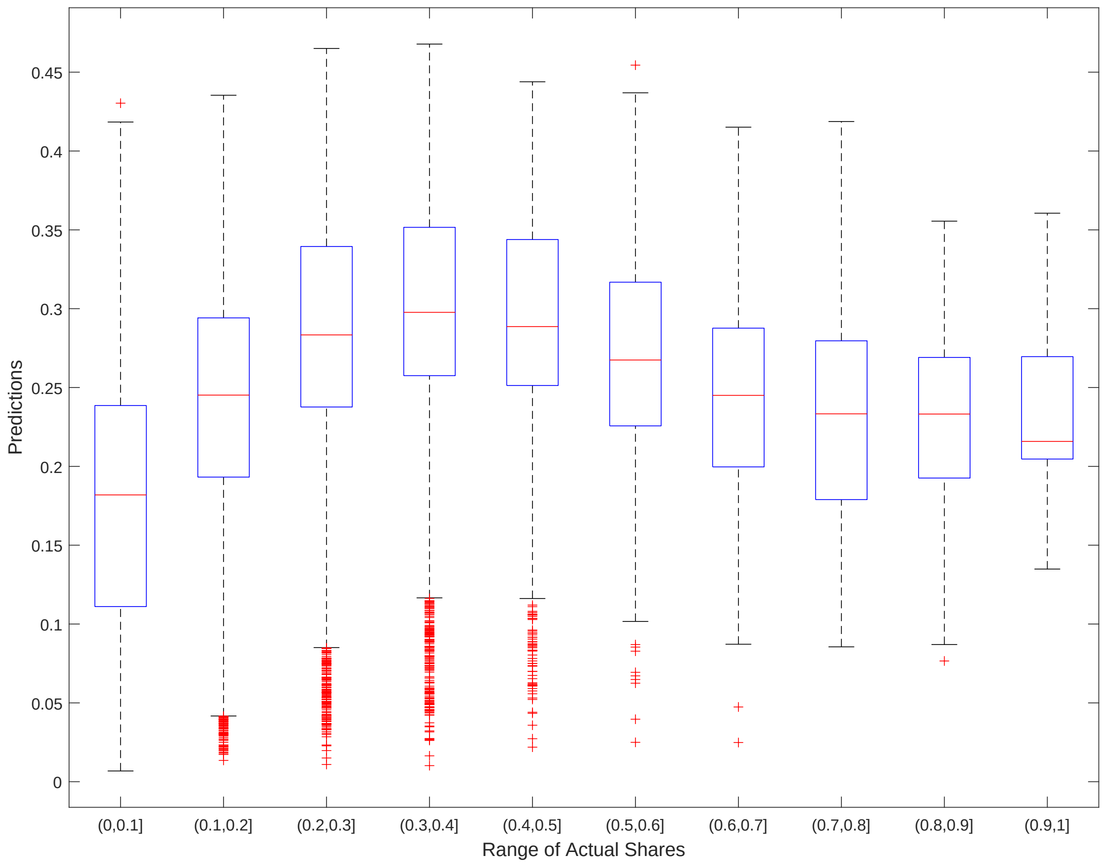

Figure 3.17: Other predictions box-plot, SUR-HEAR

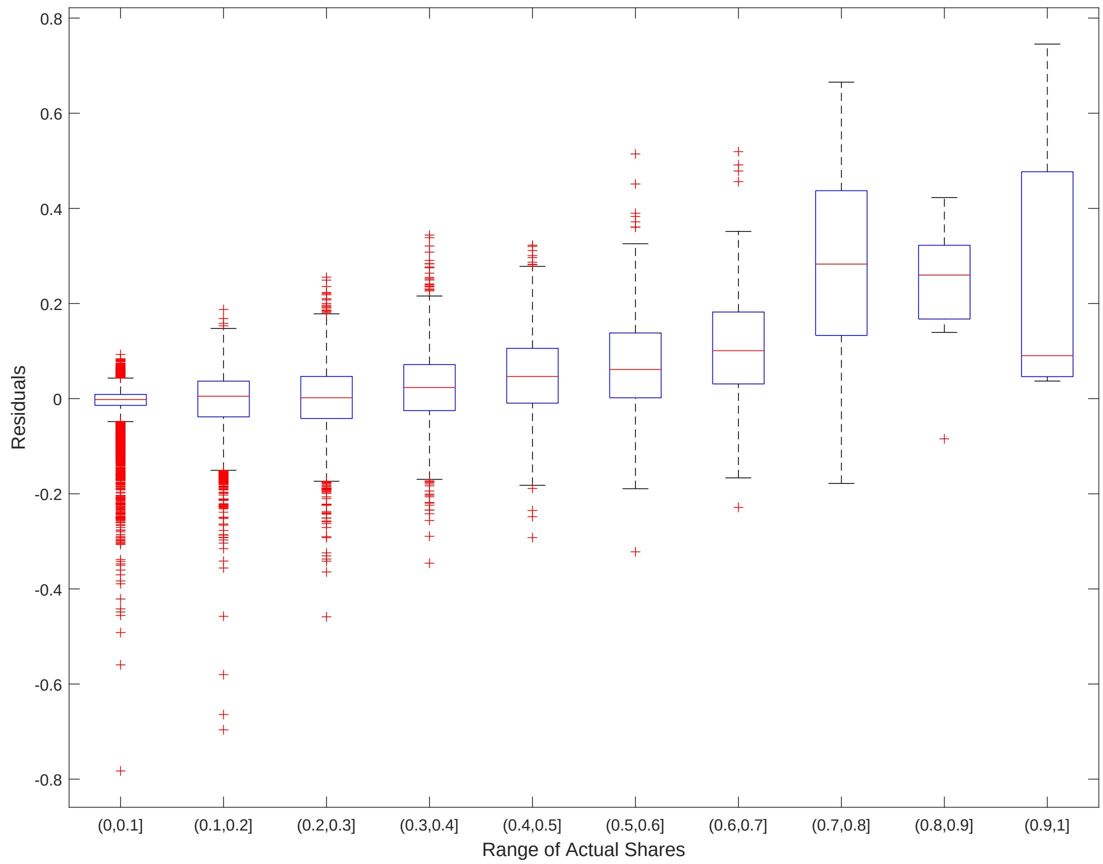

Figure 3.18: Corn residuals box-plot, best ANN

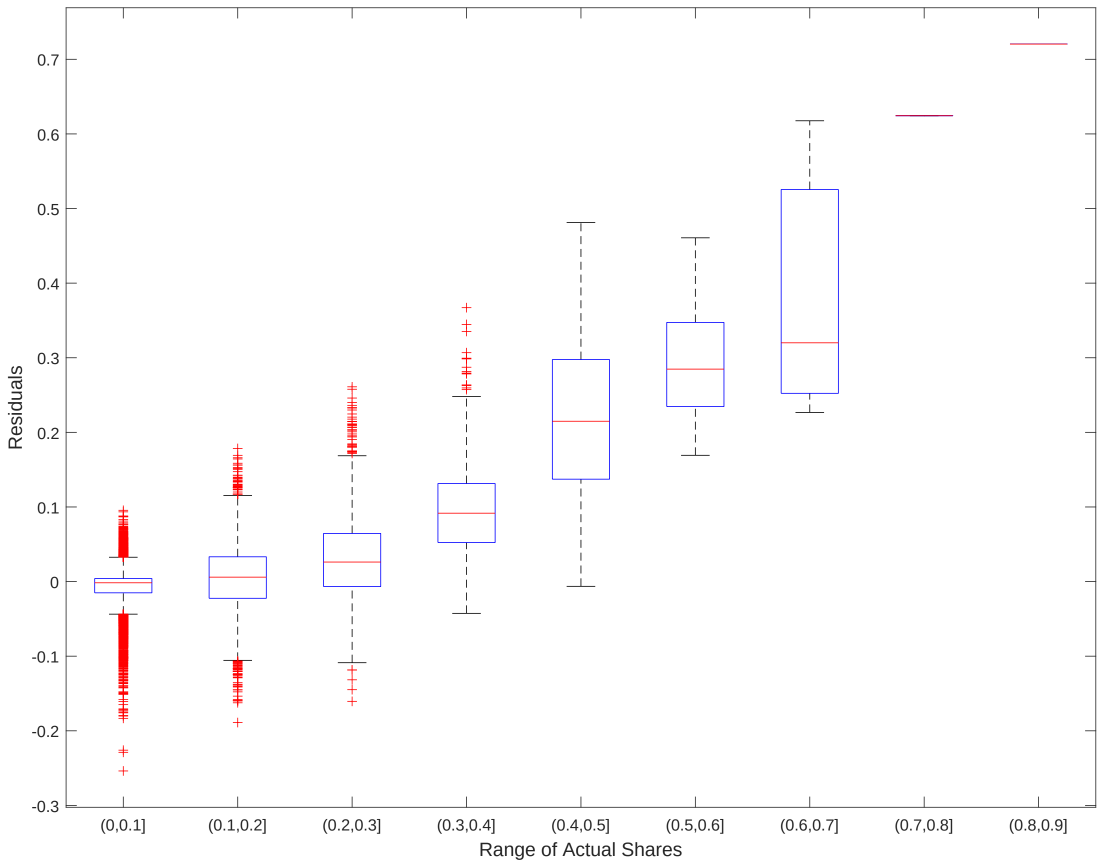

Figure 3.19: Sorghum residuals box-plot, best ANN

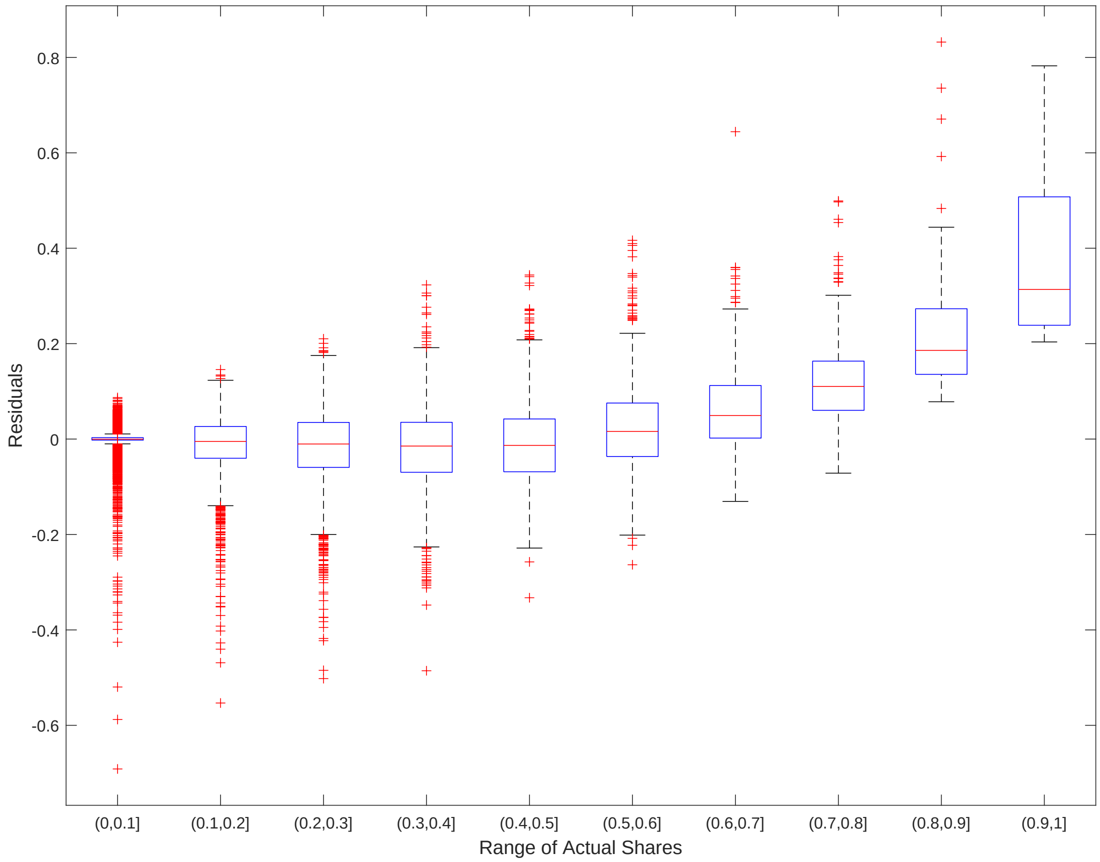

Figure 3.20: Soybean residuals box-plot, best ANN

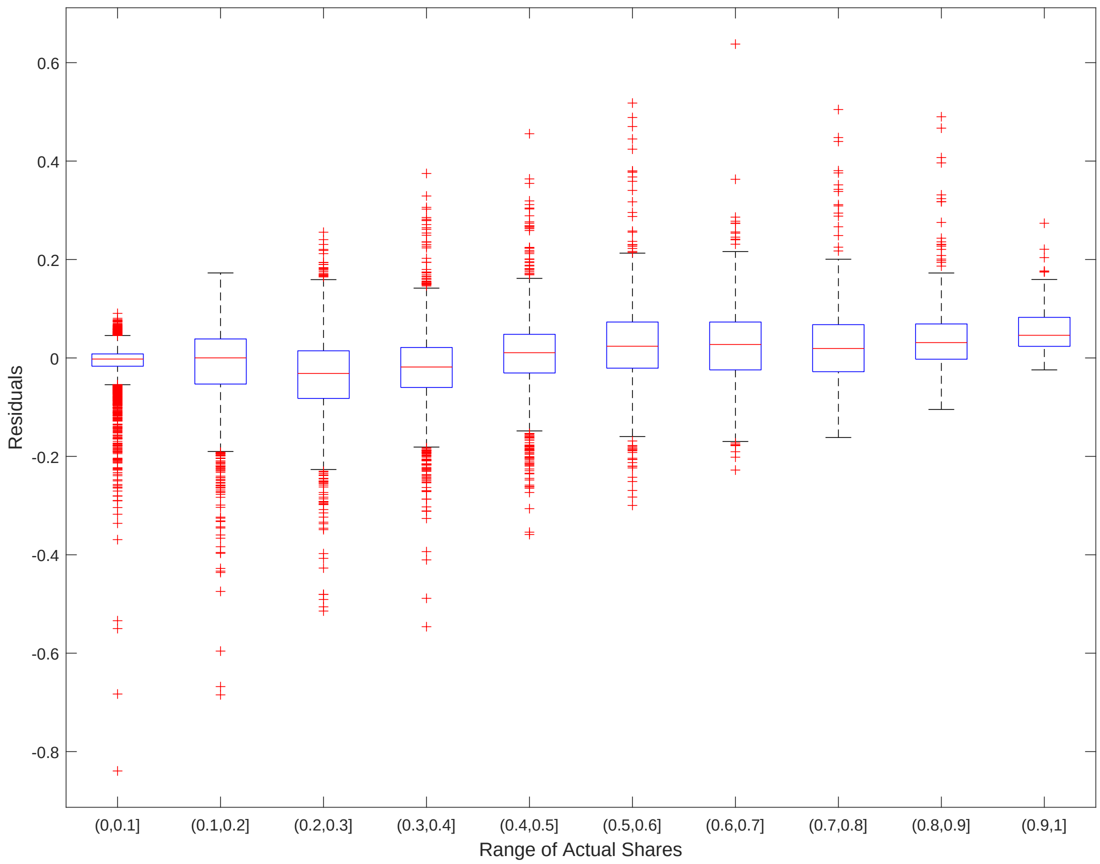

Figure 3.21: Wheat residuals box-plot, best ANN

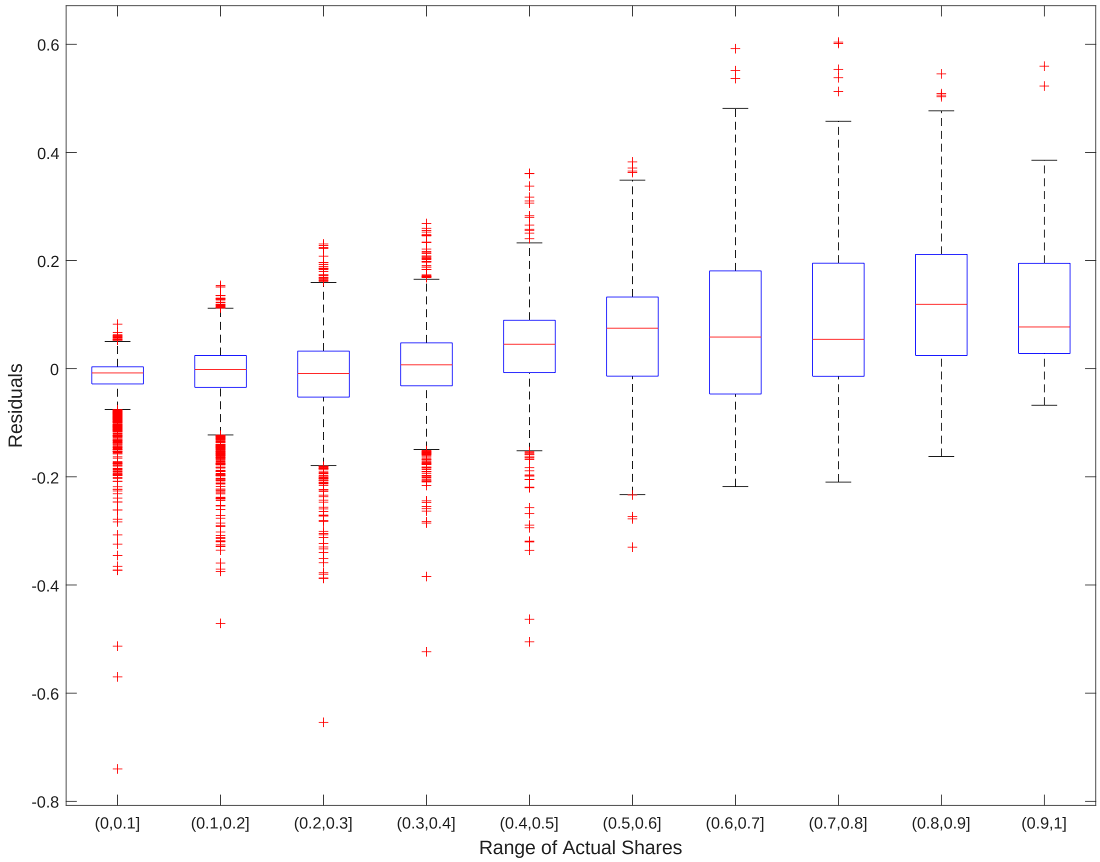

Figure 3.22: Other residuals box-plot, best ANN

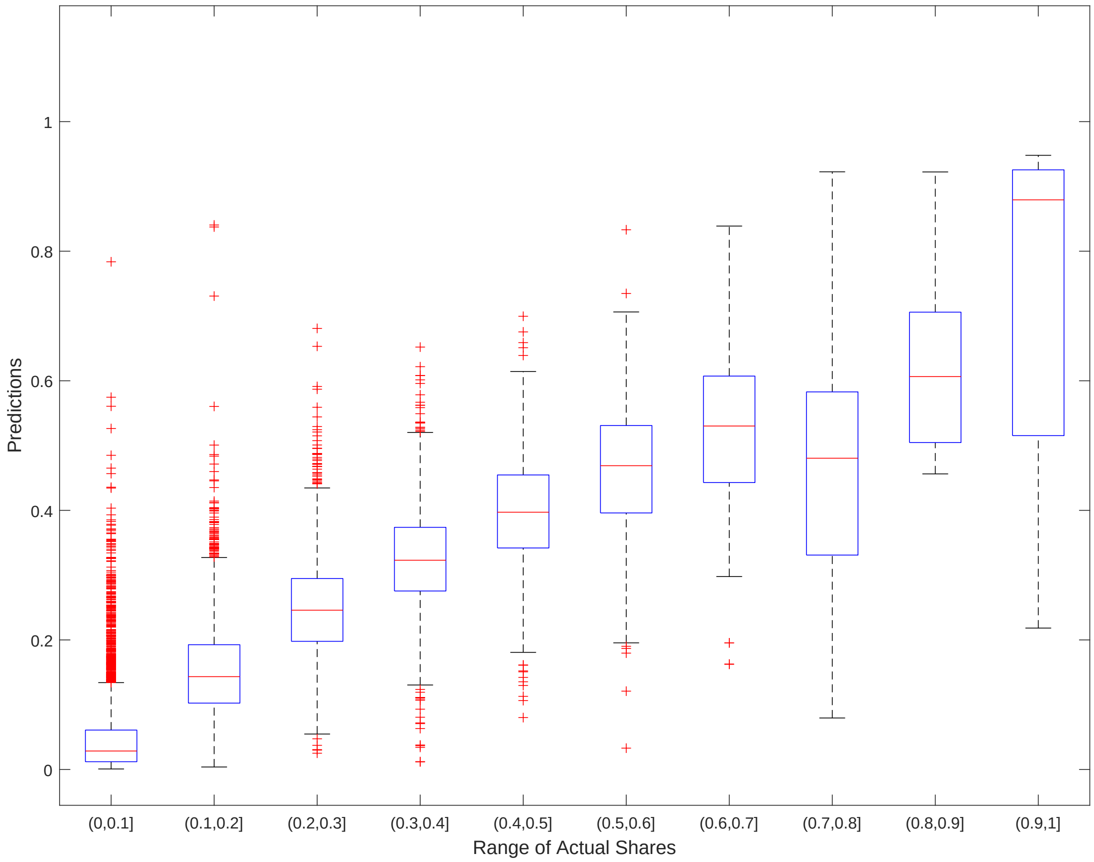

Figure 3.23: Corn predictions box-plot, best ANN

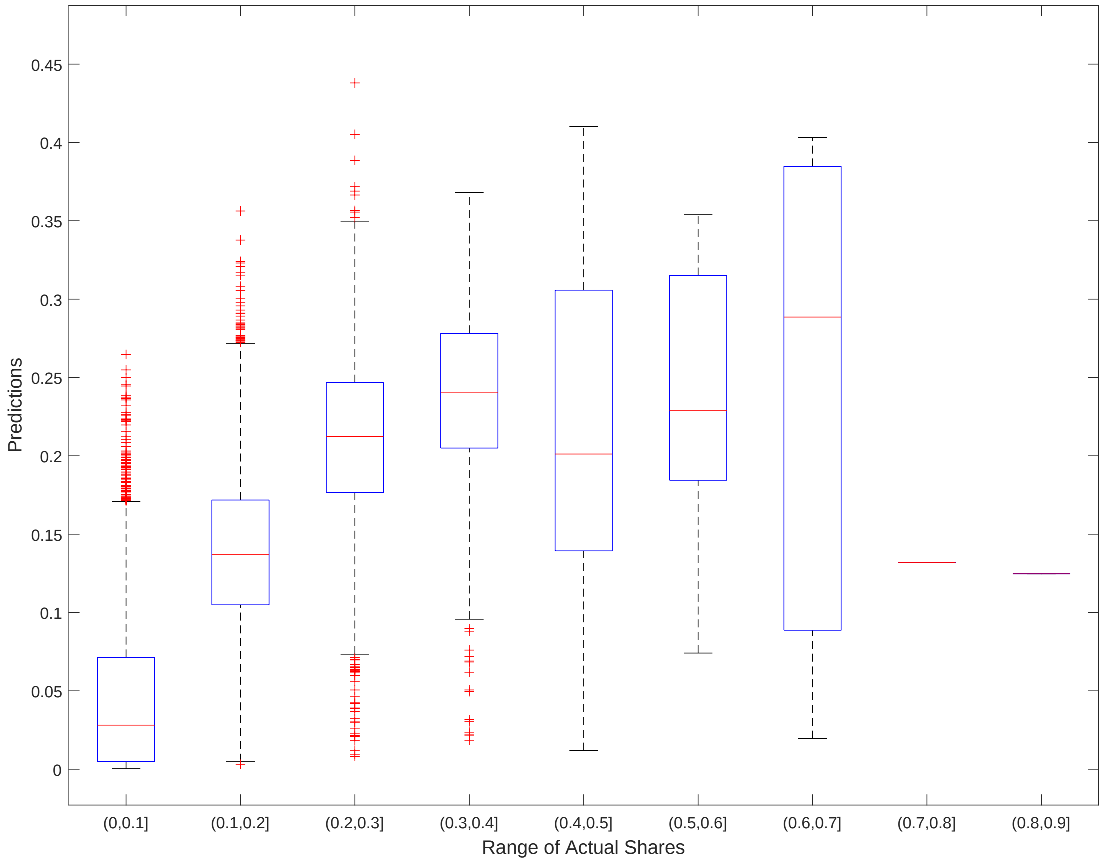

Figure 3.24: Sorghum predictions box-plot, best ANN

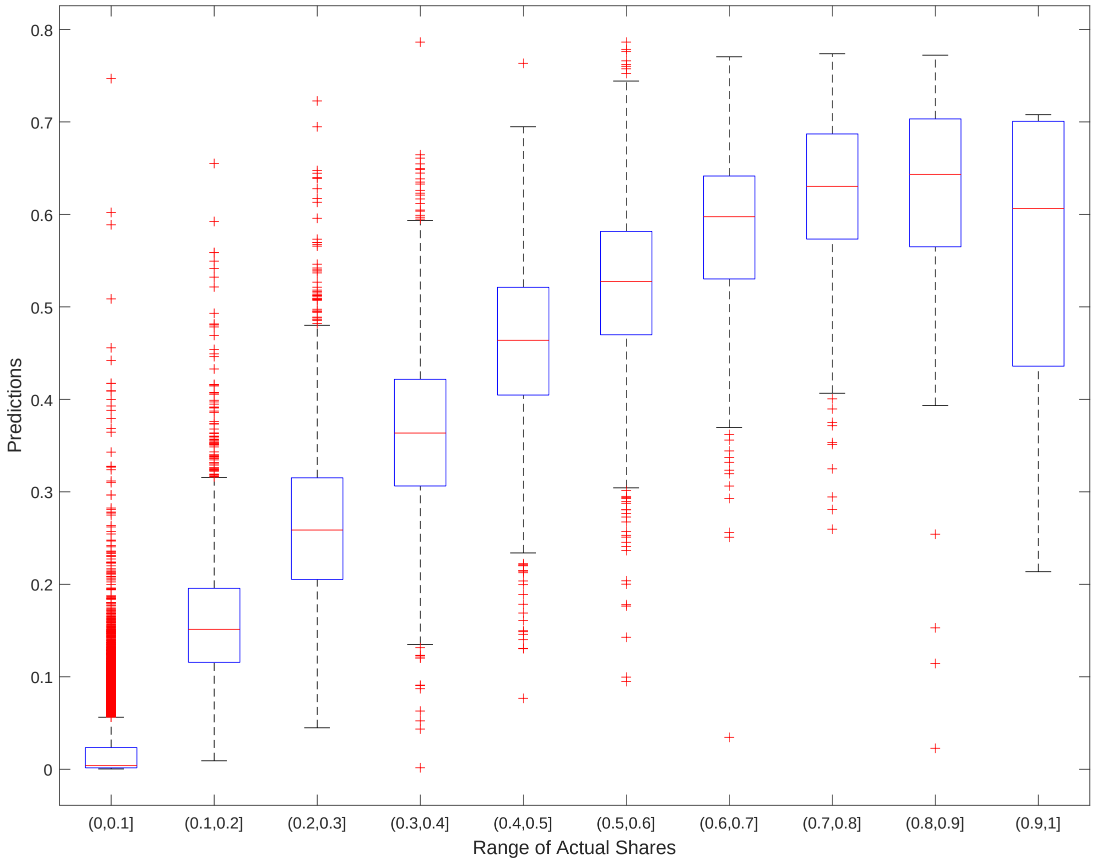

Figure 3.25: Soybeans predictions box-plot, best ANN

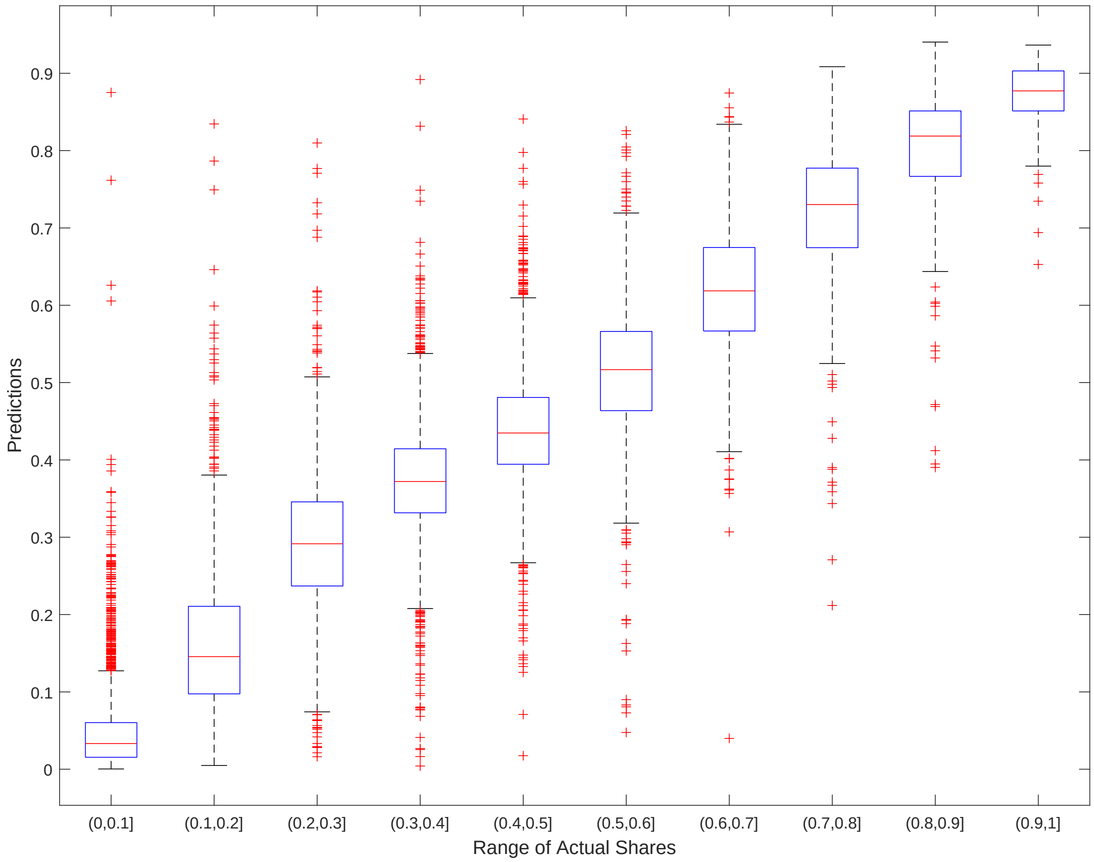

Figure 3.26: Wheat predictions box-plot, best ANN

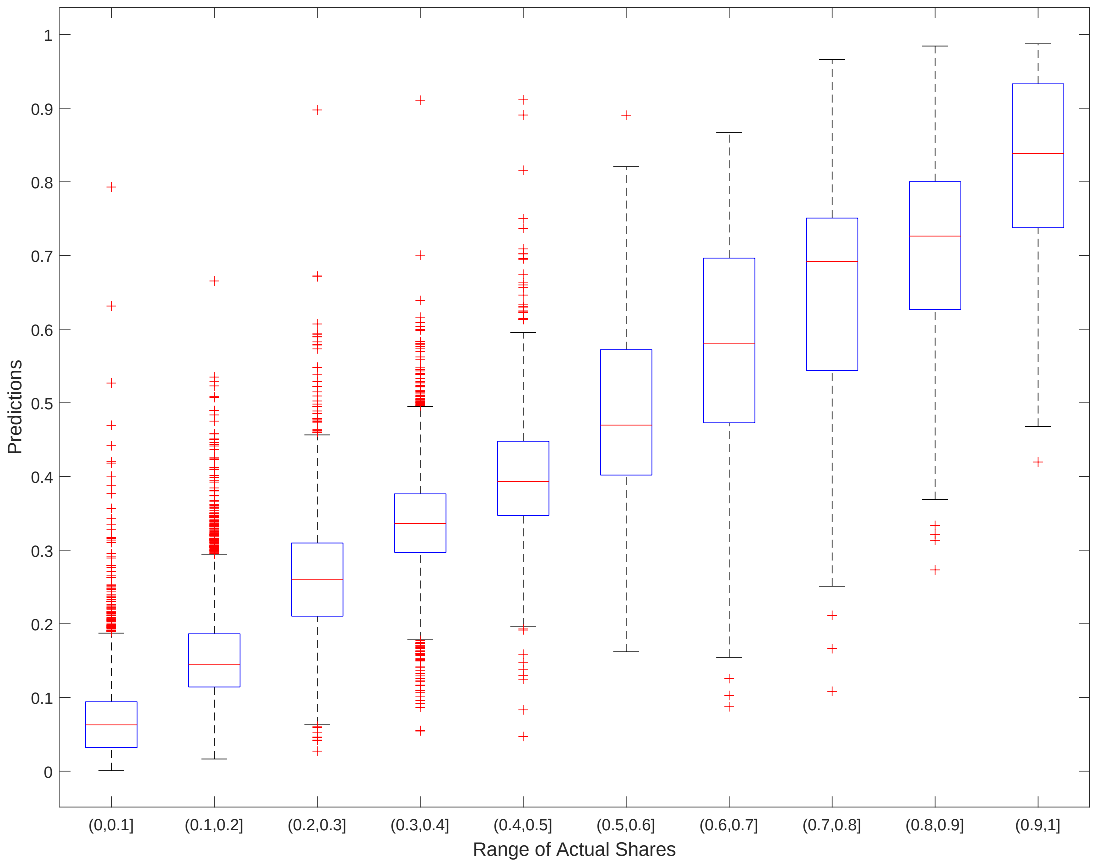

Figure 3.27: Other predictions box-plot, best ANN

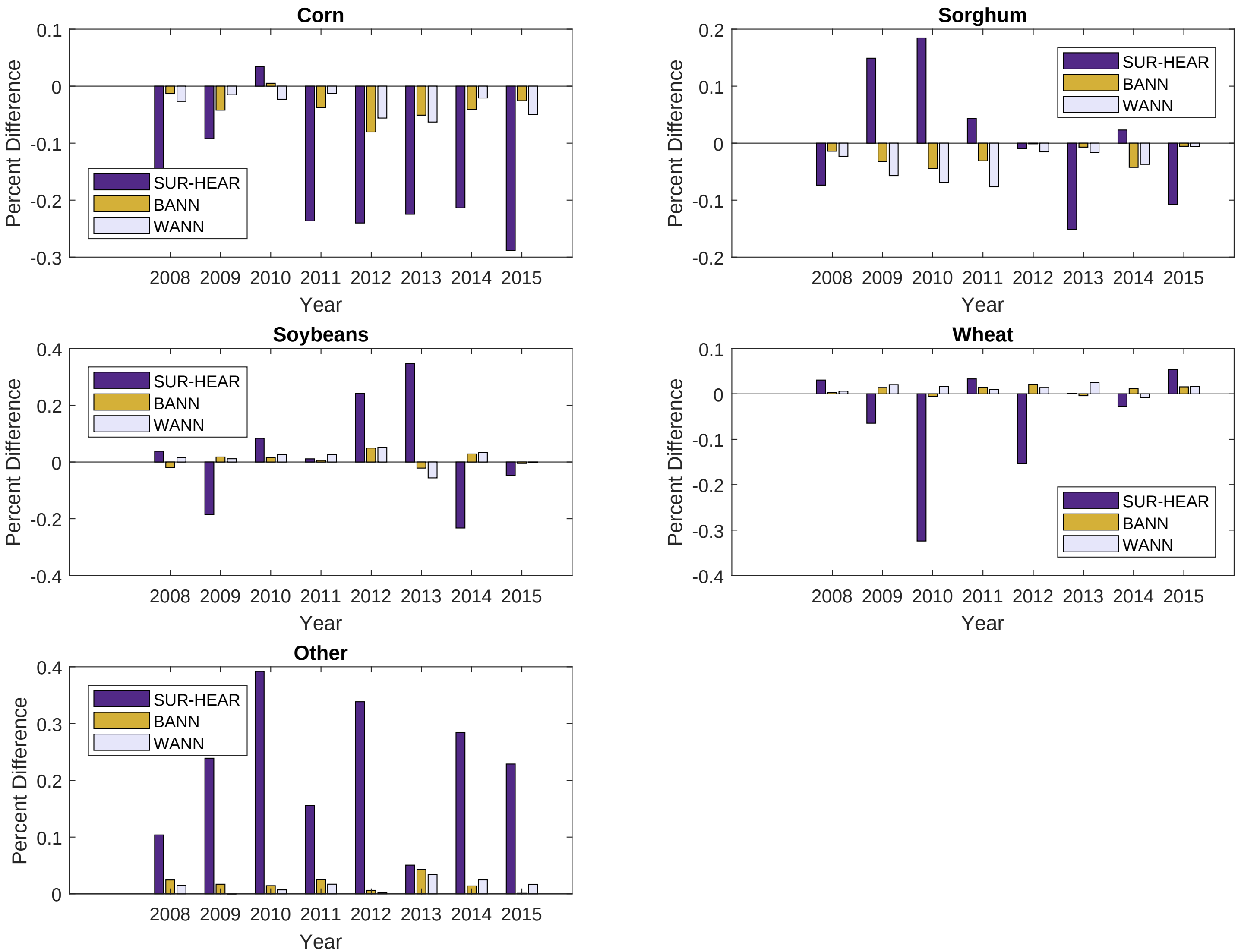

Figure 3.28: Total acre predictions: Percent deviation from actual as decimal

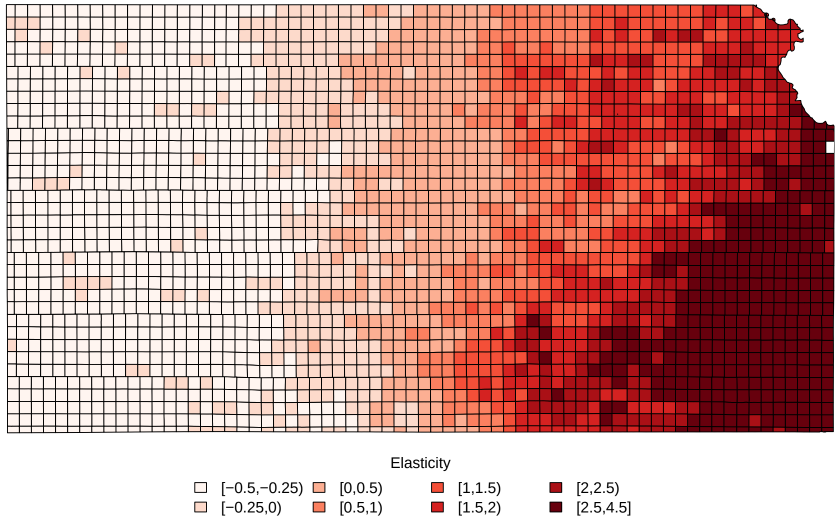

Figure 3.29: Corn own-price township-level elasticities, SUR-HEAR

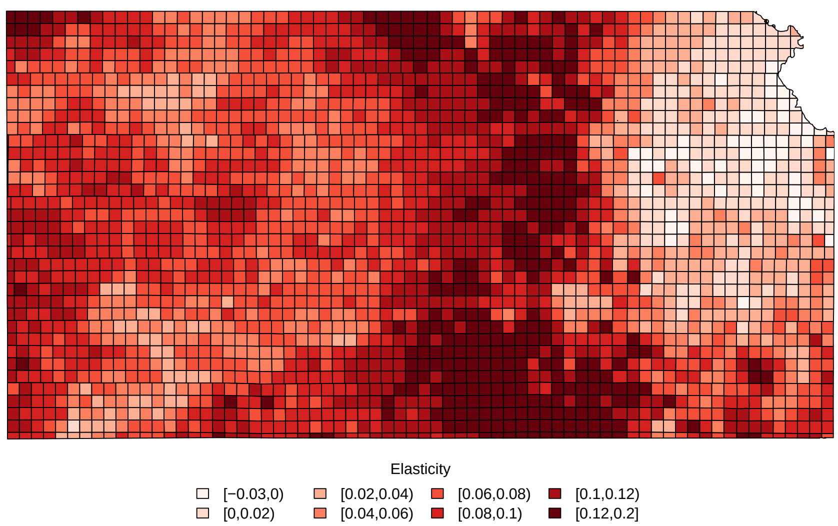

Figure 3.30: Corn own-price township-level elasticities, ANN

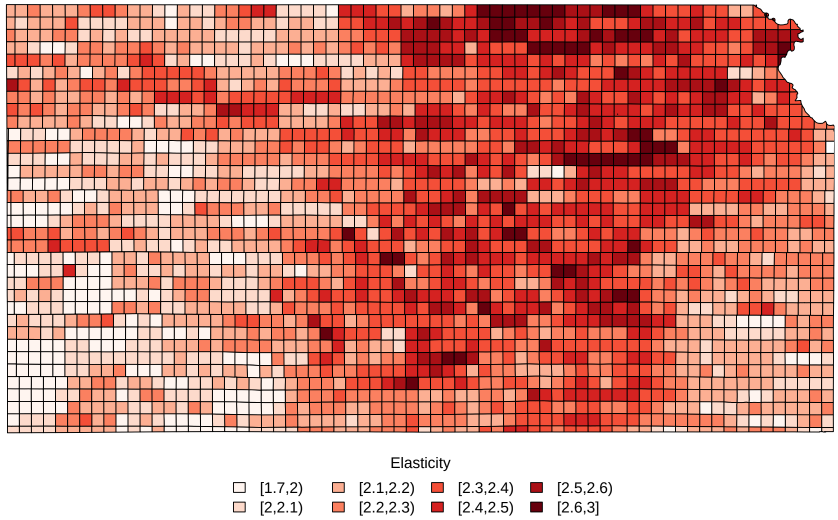

Figure 3.31: Sorghum own-price township-level elasticities, SUR-HEAR

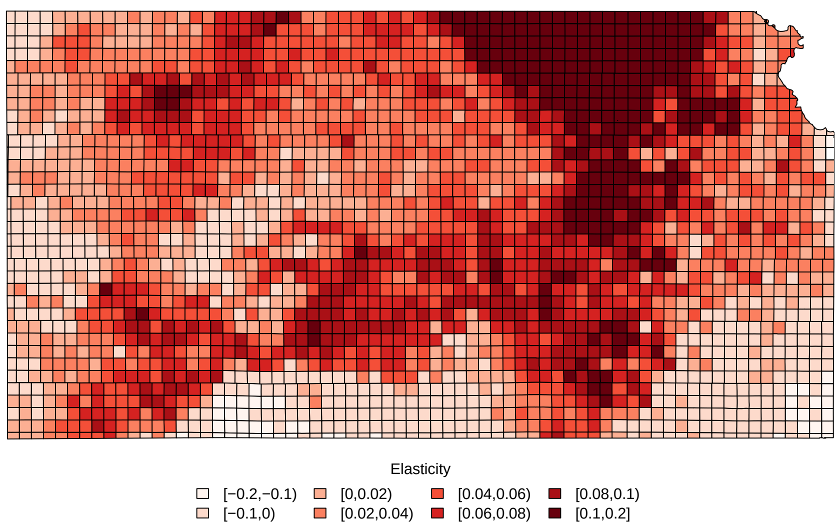

Figure 3.32: Sorghum own-price township-level elasticities, ANN

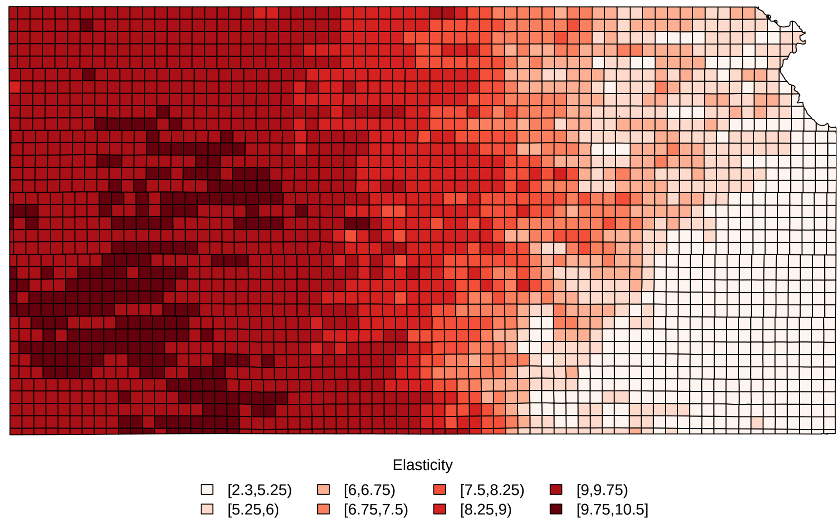

Figure 3.33: Soybeans own-price township-level elasticities, SUR-HEAR

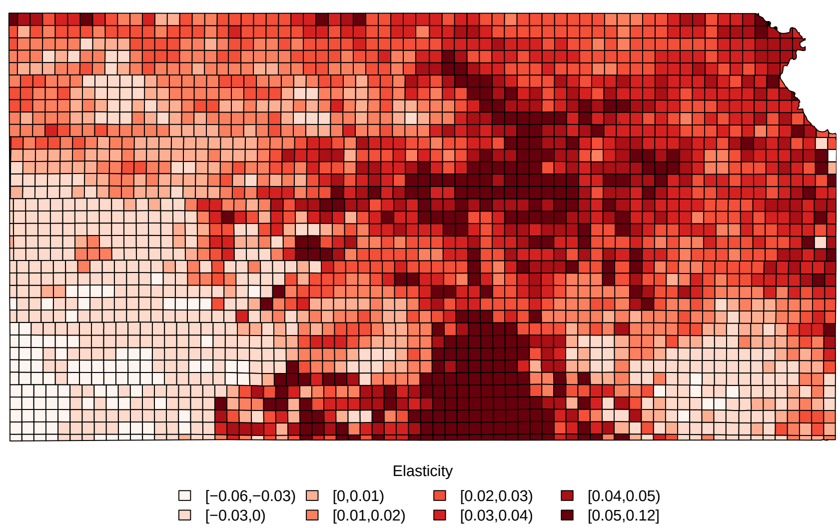

Figure 3.34: Soybeans own-price township-level elasticities, ANN

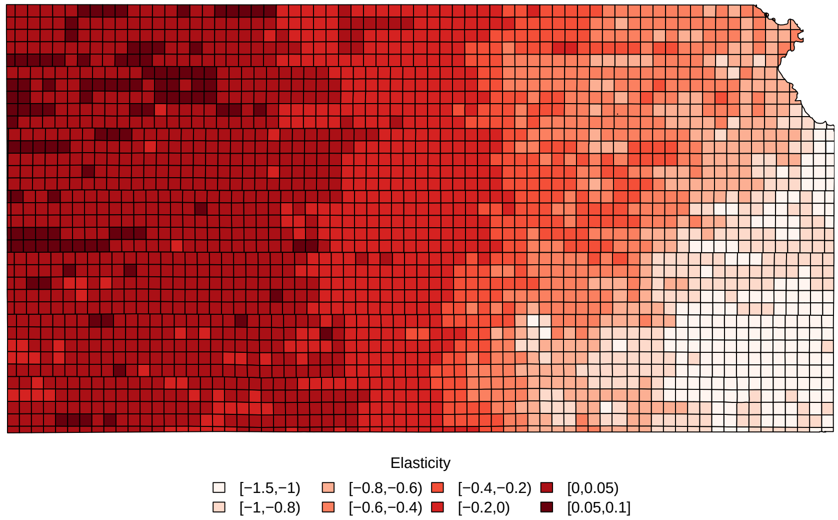

Figure 3.35: Wheat own-price township-level elasticities, SUR-HEAR

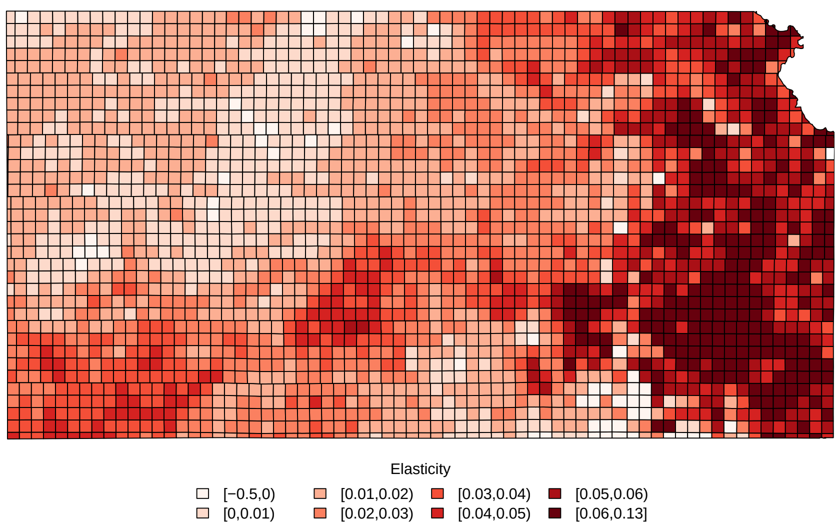

Figure 3.36: Wheat own-price township-level elasticities, ANN

Tables

Table 3.1: *Acreage elasticity estimates from previous studies*

Study	Area	Time	Corn	Sorghum	Soybeans	Wheat
Adusumilli, Rister, Lacewell, et al. (2011)	Texas	1999-2009	0.06	-0.33	0.14	-0.12
Arnade and Kelch (2007)	Iowa	1960-1999	0.01		0.48	
Bailey and Womack (1985)	CO, KS, NE, NM, OK, TX, WY	1962-1981				0.25
Bridges and Tenkorang (2009)	NE, IL, IN, IA	1986-2007	0.15 to 0.22		0.22 to 0.90	-0.92 to -0.86
Chavas and Holt (1990)	United States	1954-1985	0.17		0.45	
Chembezi and Womack (1992)	Corn Belt, Lake States, Northern High Plains	1966-1989	0.16			0.11
Hendricks et al. (2014)	IA, IL, IN	1999-2010	0.40		0.36	
Huang, Khanna, et al. (2010)	United States	1997-2007	0.51		0.49	0.07
Lee and Helmberger (1985)	IL, IND, IA, OH	1948-1980	0.12 to 0.25		0.02 to 0.35	
W. Lin and Dismukes (2007)	North Central U.S.	1991-2001	0.17 to 0.35		0.30	0.25 to 0.34
McIntosh and Shideed (1989)	Iowa	1957-1982	0.02 to 0.19			
Miao, Khanna, and Huang (2016)	United States	1997-2007	0.45		0.63	
Orazem and Miranowski (1994)	Iowa	1952-1991	0.10		0.33 to 0.38	
Wu, Mapp, and Bernardo (1996)	CO, KS, NE, NM, OK, TX, WY	1972-1988	0.05 to 0.54	0.02 to 0.80		0.03 to 0.22
Wu and Adams (2002)	Corn Belt	1982-1992	0.03 to 0.25		0.06 to 0.24	

Table 3.2: *Summary data for land use shares*

	Average Shares				
Year	Corn	Sorghum	Soybeans	Wheat	Other
2007	0.176	0.107	0.136	0.416	0.166
2008	0.175	0.108	0.158	0.372	0.187
2009	0.180	0.105	0.196	0.340	0.180
2010	0.209	0.084	0.237	0.302	0.167
2011	0.212	0.089	0.203	0.312	0.183
2012	0.196	0.088	0.193	0.324	0.199
2013	0.144	0.112	0.115	0.317	0.312
2014	0.152	0.103	0.158	0.328	0.260
2015	0.162	0.123	0.168	0.317	0.230

Table 3.3: *Approximate acreages and classifications for Cropland Data Layer codes*

Code	Description	2007	2008	2009	2010	2011	2012	2013	2014	2015	Classification
0	Background	1798	11	11	0	0	0	0	0	0	Non-Ag
1	Corn	3412343	3742110	3756324	4523569	4666376	4379140	4130037	3554649	3952194	Corn
2	Cotton	18654	10630	14576	19158	34942	24527	12100	12513	9541	Other
4	Sorghum	2195218	2721115	2586812	2177502	2322492	2243759	3054562	2444723	3280108	Sorghum
5	Soybeans	1696888	2592446	2839125	3641111	3296342	3014427	2855976	2874318	3054985	Soybeans
6	Sunflower	34305	49985	46210	47025	35564	20192	14067	10834	24516	Other
10	Peanuts	0	0	0	63	0	0	0	0	0	Other
12	Sweet Corn	0	12	0	0	55	37	114	0	0	Other
13	Pop or Orn Corn	0	0	0	0	4	0	0	37	146	Other
21	Barley	1666	1022	1050	1361	340	625	1619	2029	2983	Other
22	Durum Wheat	0	0	0	0	0	51	0	0	0	Other
23	Spring Wheat	655	0	234	0	214	427	116	0	234	Other
24	Winter Wheat	8538901	9152583	8549423	7855987	7922804	8262274	8545842	7776777	8316434	Wheat
25	Other Small Grains	9655	10057	8951	0	0	0	0	0	0	Other
26	Dbl Crop WinWht/Soybeans	407661	620000	472924	352969	472291	712648	593050	528576	573196	Other
27	Rye	19260	29223	25567	27084	30492	26861	39874	33288	49405	Other
28	Oats	6024	5400	9059	7813	7322	27390	38168	16072	35256	Other
29	Millet	3876	1438	0	3219	6730	1423	17130	1618	5773	Other
30	Speltz	0	0	0	0	0	6	0	0	0	Other
31	Canola	140	2415	3420	5626	8043	15395	19353	43968	29049	Other
33	Safflower	0	0	384	2274	653	737	145	0	0	Other
36	Alfalfa	509638	640088	558491	586120	473799	549076	551114	433226	523458	Other
37	Other Hay/Non Alfalfa	0	0	0	1	19	4	2931219	1151267	400336	Other
39	Buckwheat	0	0	0	0	0	0	0	34	0	Other
42	Dry Beans	0	0	624	42	188	109	65	1605	0	Other
43	Potatoes	1608	2482	2719	2715	3355	3537	3089	2603	1244	Other
44	Other Crops	1	1102	524	449	635	334	3258	1347	2325	Other

Table 3.3: *(continued)*

Code	Description	2007	2008	2009	2010	2011	2012	2013	2014	2015	Classification
47	Misc Vegs & Fruits	22	2	0	0	0	0	0	0	0	Other
48	Watermelons	1	0	0	0	0	0	0	0	0	Other
53	Peas	267	75	521	1183	5430	11771	3392	2493	3142	Other
57	Herbs	0	0	0	155	60	123	81	0	0	Other
58	Clover/Wildflowers	577	6011	1428	1294	1393	866	3522	2612	1849	Other
59	Sod/Grass Seed	0	2431	1739	0	0	0	1882	384	887	Other
60	Switchgrass	0	0	0	0	0	0	1175	0	0	Other
61	Fallow/Idle Cropland	2494602	3107615	3190824	3108558	3276103	3216495	3215190	3153360	3337227	Other
63	Forest	171	412	140	0	0	0	0	0	0	Other
66	Cherries	0	0	0	0	0	0	0	14	0	Other
68	Apples	6	12	0	0	6	11	0	0	0	Other
69	Grapes	0	0	0	0	26	4	0	0	0	Other
70	Christmas Trees	0	0	0	0	12	0	0	0	0	Other
76	Walnuts	0	0	0	0	62	33	0	0	0	Other
71	Other Tree Crops	76	83	45	0	14	0	0	0	0	Other
74	Pecans	0	0	0	28	6	7	800	101	83	Other
87	Wetlands	2777	1053	1025	0	0	0	0	0	0	Non-Ag
111	Open Water	442795	570122	433436	444605	412277	391713	401175	464814	454111	Non-Ag
121	Developed/Open Space	2830961	3049334	2804152	1947722	1974857	1968895	1956278	1997921	1970304	Non-Ag
122	Developed/Low Intensity	533040	529052	536309	531099	546445	550665	541284	555101	551462	Non-Ag
123	Developed/Med Intensity	102536	103461	102814	113071	122079	122820	122946	148689	149994	Non-Ag
124	Developed/High Intensity	37501	40569	38728	42314	45299	45354	44054	52417	50621	Non-Ag
131	Barren	16916	25972	12841	15505	18503	16939	29397	29173	24916	Non-Ag
141	Deciduous Forest	2439431	3390542	2530175	2564152	2345042	2326176	2267055	2258277	2528637	Non-Ag
142	Evergreen Forest	1729	3263	3263	3323	4210	5018	3654	3628	4074	Non-Ag
143	Mixed Forest	5322	13300	9576	7067	12006	14304	13432	10919	10046	Non-Ag
152	Shrubland	118473	155896	147086	163437	139895	139342	171644	163625	199352	Non-Ag

Table 3.3: *(continued)*

Code	Description	2007	2008	2009	2010	2011	2012	2013	2014	2015	Classification
176	Grassland/Pasture	26547194	21754945	23729246	24081434	24058519	24194289	20599410	24481004	22581186	Non-Ag
190	Woody Wetlands	263781	350720	272510	286564	272198	257351	293638	288534	344352	Non-Ag
195	Herbaceous Wetlands	7455	16668	11372	9558	17601	17399	14324	20672	16065	Non-Ag
205	Triticale	0	0	0	7955	8857	10023	24035	27652	41137	Other
224	Vetch	0	0	0	28	88	123	34	0	0	Other
225	Dbl Crop WinWht/Corn	0	0	0	12248	19307	11639	15236	5611	10286	Other
226	Dbl Crop Oats/Corn	0	0	0	13	0	6	482	51	220	Other
229	Pumpkins	0	0	0	1	11	4	31	0	12	Other
235	Dbl Crop Barley/Sorghum	0	0	0	103	204	132	674	398	0	Other
236	Dbl Crop WinWht/Sorghum	0	0	0	107705	140288	118194	163987	143348	158860	Other
237	Dbl Crop Barley/Corn	0	0	0	5	0	98	50	5	0	Other
238	Dbl Crop WinWht/Cotton	0	0	0	0	22	37	63	137	80	Other
239	Dbl Crop Soybeans/Cotton	0	0	0	0	0	0	0	65	0	Other
240	Dbl Crop Soybeans/Oats	0	0	0	325	51	755	375	176	431	Other
241	Dbl Crop Corn/Soybeans	0	0	0	169	51	72	247	8	4	Other
246	Radishes	0	0	0	0	86	0	0	7	43	Other
247	Turnips	0	0	0	5	0	16	43	0	0	Other
254	Dbl Crop Barley/Soybeans	0	0	0	0	45	62	198	11	123	Other

Table 3.4: *Dependent and independent variables for SUR-HEAR model*

Variable	Description	mean
$CORN_t$	Log of corn share-other share	-0.49
SOR_t	Log of corn share-other share	-1.14
SOY_t	Log of corn share-other share	-1.38
WHT_t	Log of corn share-other share	0.38
$PCORN$	Expected corn price	4.68
$PSOR$	Expected sorghum price	4.37
$PSOY$	Expected soybean price	10.40
$PWHEAT$	Expected wheat price	6.30
$PDIESEL$	Diesel price index	83.87
$PLABOR$	Labor price index	100.80
$CLAY$	Percent clay in soil	0.24
$SILT$	Percent silt in soil	0.46
$TFACTOR$	T-factor: maximum sustainable erosion (tons/ac/year)	4.49
WEI	Wind erosion index	69.37
AVG_{CSS}	3-year-average total growing-season precipitation (Apr-Aug) (mm), corn/sorghum/soybeans.	356.87
AVG_W	3-year-average total growing-season precipitation (Nov-Jun) (mm), wheat.	435.78
$PREC_{CSS}$	Total planting season (May-Jun) precipitation (mm), corn/sorghum/soybeans.	295.39
$PREC_W$	Total planting season (Sep-Oct) precipitation (mm), wheat.	133.89
$TMAX_{CSS}$	Average-daily planting-season (May-Jun) maximum temperature (C°), corn/sorghum/soybeans,	24.84
$TMAX_W$	Average-daily planting-season (Sep-Oct) maximum temperature (C°), wheat	24.05
$TMIN_{CSS}$	Average-daily planting-season (May-Jun) minimum temperature (C°), corn/sorghum/soybeans	10.70
$TMIN_W$	Average-daily planting-season (Sep-Oct) minimum temperature (C°), wheat	9.09

$N = 21,078$

Table 3.5: *Dependent and independent variables for ANN models*

Variable	Description	mean
$CORN_t$	Log of corn share-other share	-0.51
SOR_t	Log of sorghum share-other share	-1.21
SOY_t	Log of soybean share-other share	-1.35
WHT_t	Log of wheat share-other share	0.29
$CORN_{t-1}$	Lag of $CORN_t$	-0.44
SOR_{t-1}	Lag of SOR_t	-1.14
SOY_{t-1}	Lag of SOY_t	-1.33
WHT_{t-1}	Lag of WHT_t	0.42
$PCORN$	Expected corn price	4.79
$PSOR$	Expected sorghum price	4.46
$PSOY$	Expected soybean price	10.79
$PWHEAT$	Expected wheat price	6.54
$PDIESEL$	Diesel price index	86.20
$PLABOR$	Labor price index	101.96
$CLAY$	Percent clay in soil	0.24
$SILT$	Percent silt in soil	0.46
$TFACTOR$	T-factor: maximum sustainable erosion (tons/ac/year)	4.49
WEI	WEI: wind erosion index	69.37
AVG_{CSS}	3-year-average total growing-season precipitation (Apr-Aug) (mm), corn/sorghum/soybeans.	350.83
AVG_W	3-year-average total growing-season precipitation (Nov-Jun) (mm), wheat.	435.44
$PREC_{CSS}$	Total planting season (May-Jun) precipitation (mm), corn/sorghum/soybeans.	284.64
$PREC_W$	Total planting season (Sep-Oct) precipitation (mm), wheat.	137.07
$TMAX_{CSS}$	Average-daily planting-season (May-Jun) maximum temperature (C°), corn/sorghum/soybeans,	25.03
$TMAX_W$	Average-daily planting-season (Sep-Oct) maximum temperature (C°), wheat	24.22
$TMIN_{CSS}$	Average-daily planting-season (May-Jun) minimum temperature (C°), corn/sorghum/soybeans	10.69
$TMIN_W$	Average-daily planting-season (Sep-Oct) minimum temperature (C°), wheat	9.22

$N = 18,736$

Table 3.6: *SUR-HEAR test statistics*

Equation	Test Statistic	Critical Value	p-Value	Decision
Autocorrelation, H_0: No autocorrelation				
Corn	4.3843	3.8415	0.05	Reject H_0
Sorghum	3.3323	3.1701	0.075	Reject H_0
Soybeans	3.507	3.1701	0.075	Reject H_0
Wheat	3.7622	3.1701	0.075	Reject H_0
Heteroskedasticity, H_0: No heteroskedasticity				
Corn	12,705	2,503	0.01	Reject H_0
Sorghum	10,512	2,503	0.01	Reject H_0
Soybeans	7,728	2,503	0.01	Reject H_0
Wheat	16,510	2,503	0.01	Reject H_0
Contemporaneous correlation, H_0: No contemporaneous correlation				
All	2,297	16.81	0.01	Reject H_0

Table 3.7: *Results for Ramsey RESET test*

Equation	Variable	Coefficient p-Value
Corn	$\hat{CORN}_t^2$	0.000
Corn	$\hat{CORN}_t^3$	0.000
Sorghum	$\hat{SOR}_t^2$	0.000
Sorghum	$\hat{SOR}_t^3$	0.001
Soybeans	$\hat{SOY}_t^2$	0.000
Soybeans	$\hat{SOY}_t^3$	0.000
Wheat	$\hat{WHT}_t^2$	0.000
Wheat	$\hat{WHT}_t^3$	0.000

Table 3.8: *Mean-square error results, SUR-HEAR*

Year	Corn	Sorghum	Soybeans	Wheat	Other	All
2007	0.027	0.007	0.017	0.042	0.015	0.022
2008	0.026	0.007	0.019	0.036	0.015	0.020
2009	0.028	0.007	0.023	0.033	0.016	0.021
2010	0.033	0.005	0.019	0.039	0.013	0.022
2011	0.033	0.006	0.018	0.026	0.014	0.019
2012	0.028	0.005	0.023	0.029	0.018	0.021
2013	0.019	0.006	0.017	0.027	0.051	0.024
2014	0.018	0.007	0.020	0.033	0.036	0.023
2015	0.022	0.006	0.024	0.027	0.020	0.020
All	0.026	0.006	0.020	0.032	0.022	0.021

Table 3.9: *Average-mean-square error results, ANNs*

Year	Corn	Sorghum	Soybeans	Wheat	Other	All
2008	0.006	0.002	0.004	0.006	0.003	0.004
2009	0.005	0.002	0.005	0.006	0.004	0.004
2010	0.005	0.002	0.005	0.005	0.003	0.004
2011	0.006	0.002	0.006	0.006	0.004	0.005
2012	0.006	0.002	0.006	0.006	0.005	0.005
2013	0.004	0.003	0.003	0.005	0.010	0.005
2014	0.004	0.003	0.004	0.006	0.005	0.004
2015	0.005	0.003	0.006	0.005	0.005	0.005
All	0.005	0.002	0.005	0.006	0.005	0.005

Table 3.10: *Deviations from actual acreages, SUR-HEAR*

Year	Corn	Sorghum	Soybeans	Wheat	Other
2007	-0.15	0.15	-0.01	-0.04	0.15
2008	-0.17	-0.07	0.04	0.03	0.10
2009	-0.09	0.15	-0.18	-0.06	0.24
2010	0.03	0.18	0.08	-0.32	0.39
2011	-0.24	0.04	0.01	0.03	0.16
2012	-0.24	-0.01	0.24	-0.15	0.34
2013	-0.22	-0.15	0.35	0.00	0.05
2014	-0.21	0.02	-0.23	-0.03	0.28
2015	-0.29	-0.11	-0.05	0.05	0.23
MAD	0.18	0.10	0.13	0.08	0.22

Table 3.11: *Deviations from actual acreages, best and worst ANNs*

	Corn		Sorghum		Soybeans		Wheat		Other	
YEAR	BANN	WANN	BANN	WANN	BANN	WANN	BANN	WANN	BANN	WANN
2008	-0.01	-0.03	-0.01	-0.02	-0.02	0.02	0.00	0.01	0.02	0.01
2009	-0.04	-0.02	-0.03	-0.06	0.02	0.01	0.01	0.02	0.02	0.00
2010	0.01	-0.02	-0.04	-0.07	0.02	0.03	-0.01	0.02	0.01	0.01
2011	-0.04	-0.01	-0.03	-0.08	0.01	0.03	0.01	0.01	0.02	0.02
2012	-0.08	-0.06	0.00	-0.02	0.05	0.05	0.02	0.01	0.01	0.00
2013	-0.05	-0.06	-0.01	-0.02	-0.02	-0.06	0.00	0.02	0.04	0.03
2014	-0.04	-0.02	-0.04	-0.04	0.03	0.03	0.01	-0.01	0.01	0.02
2015	-0.03	-0.05	-0.01	-0.01	0.00	0.00	0.02	0.02	0.00	0.02
MAD	0.04	0.03	0.02	0.04	0.02	0.03	0.01	0.01	0.02	0.01

Table 3.12: *SUR-HEAR and ANN average-annual aggregate elasticities*

	SUR-HEAR				ANNs			
Variable	Corn	Sorghum	Soybeans	Wheat	Corn	Sorghum	Soybeans	Wheat
$CORN_{t-1}$					0.01^{***} (6.38)	0.00 (−1.21)	-0.01^{***} (−3.33)	0.00^{**} (−2.06)
SOR_{t-1}					0.01^{***} (4.49)	-0.01^{***} (−14.35)	0.00 (1.07)	0.00^{*} (1.74)
SOY_{t-1}					-0.01^{***} (−3.49)	0.00 (−0.83)	0.02^{***} (4.55)	0.00 (0.00)
WHT_{t-1}					-0.01^{***} (−14.60)	-0.01^{***} (−8.27)	-0.02^{***} (−13.40)	0.02^{***} (24.88)
$PCORN$	0.66^{***} (3.65)	-0.64^{***} (−4.08)	-4.37^{***} (−16.66)	0.56^{***} (6.22)	0.06^{***} (2.86)	0.04^{*} (1.69)	−0.02 (−0.56)	−0.01 (−0.88)
$PSOR$	-2.14^{***} (−15.38)	2.15^{***} (17.12)	0.61^{***} (3.43)	0.89^{***} (11.26)	-0.03^{*} (−1.88)	0.04 (1.53)	0.01 (0.54)	0.00 (−0.34)
$PSOY$	2.08^{***} (15.55)	-2.77^{***} (−26.54)	6.68^{***} (45.82)	-2.24^{***} (−23.89)	0.00 (0.11)	−0.04 (−1.43)	0.03 (1.05)	0.01 (0.63)
$PWHEAT$	-0.70^{***} (−13.49)	0.12^{**} (2.44)	1.59^{***} (22.78)	-0.15^{***} (−5.52)	-0.04^{**} (−2.08)	0.04 (1.20)	-0.05^{**} (−2.43)	0.02 (1.28)
$PDIESEL$	0.04 (0.37)	-0.53^{***} (−5.67)	-2.26^{***} (−18.83)	0.36^{***} (6.92)	0.00 (−0.02)	0.00 (0.01)	−0.08 (−1.01)	−0.03 (−0.64)
$PLABOR$	-0.70^{**} (−2.57)	0.56^{***} (2.59)	-0.81^{***} (−3.33)	-0.22^{**} (−2.03)	−0.07 (−0.72)	0.08 (0.47)	0.05 (0.44)	−0.02 (−0.24)
AVG_{CSS}	1.11^{***} (16.03)	-1.06^{***} (−21.06)	2.93^{***} (38.99)		0.01 (1.14)	-0.02^{***} (−2.97)	0.01^{*} (1.69)	0.01^{*} (1.88)
AVG_{W}				-0.87^{***} (−19.80)	-0.02^{***} (−3.83)	0.01 (0.90)	0.01 (1.00)	0.00 (−1.15)
$PREC_{CSS}$	-0.10^{***} (−5.65)	0.04^{**} (2.11)	0.12^{***} (4.29)		0.01^{***} (3.39)	0.01^{***} (2.68)	0.00 (−1.06)	0.00^{**} (−2.16)
$PREC_{W}$				0.02^{***} (4.29)	0.00 (0.38)	0.00 (−0.49)	0.00 (0.72)	0.00 (−0.21)
$TMAX_{CSS}$	-0.57^{**} (−2.33)	4.61^{***} (16.76)	-7.28^{***} (−27.31)		0.04 (0.99)	−0.05 (−0.72)	−0.04 (−0.61)	0.03 (0.86)
$TMAX_{W}$				1.90^{***} (14.30)	−0.06 (−1.45)	0.01 (0.08)	-0.21^{***} (−4.23)	0.03 (0.98)
$TMIN_{CSS}$	-0.72^{***} (−5.66)	-1.89^{***} (−13.82)	4.18^{***} (25.37)		−0.03 (−1.64)	−0.04 (−1.26)	0.06^{*} (1.77)	0.00 (0.01)
$TMIN_{W}$				-0.49^{***} (−7.96)	0.01 (0.32)	0.02 (0.58)	0.00 (−0.13)	0.01 (0.99)

Values in parentheses denote z-statistics

***, **, * $\Rightarrow$ Significance at 1%, 5%, 10% level

Table 3.13: *SUR-HEAR bootstrapped-aggregate elasticities, corn*

	Years								
Variable	2007	2008	2009	2010	2011	2012	2013	2014	2015
$PCORN$	0.33^{**} (2.56)	0.72^{***} (4.54)	0.40^{***} (3.53)	0.59^{***} (5.54)	0.89^{***} (3.61)	0.98^{***} (4.29)	0.88^{***} (2.79)	0.49^{**} (2.13)	0.43^{***} (4.07)
$PSOR$	-1.95^{***} (-15.34)	-2.97^{***} (-15.33)	-1.63^{***} (-15.96)	-1.41^{***} (-15.07)	-2.73^{***} (-15.40)	-2.46^{***} (-14.76)	-2.39^{***} (-14.79)	-2.30^{***} (-15.53)	-1.90^{***} (-15.42)
$PSOY$	1.98^{***} (18.71)	2.72^{***} (13.77)	1.94^{***} (16.02)	1.38^{***} (10.57)	2.46^{***} (16.23)	2.03^{***} (12.85)	2.17^{***} (16.05)	2.68^{***} (17.14)	2.17^{***} (14.07)
$PWHEAT$	-0.44^{***} (-11.81)	-0.70^{***} (-14.60)	-0.68^{***} (-12.55)	-0.56^{***} (-16.13)	-0.73^{***} (-13.12)	-0.93^{***} (-14.20)	-1.07^{***} (-12.87)	-0.75^{***} (-11.96)	-0.64^{***} (-13.63)
$PDIESEL$	-0.07 (-0.63)	-0.01 (1.18)	-0.04 (0.45)	0.09^{**} (2.34)	0.08 (0.37)	0.15 (0.90)	0.07 (-0.56)	-0.11 (-1.06)	-0.03 (1.07)
$PLABOR$	-0.67^{**} (-2.22)	-0.67^{**} (-2.43)	-0.74^{***} (-2.62)	-0.69^{**} (-2.38)	-0.68^{**} (-2.52)	-0.72^{**} (-2.55)	-0.79^{***} (-2.99)	-0.88^{***} (-2.82)	-0.86^{**} (-2.45)
AVG_{CSS}	1.59^{***} (17.13)	1.37^{***} (14.66)	1.33^{***} (16.43)	0.85^{***} (10.95)	1.07^{***} (16.04)	0.97^{***} (14.14)	0.98^{***} (16.90)	1.15^{***} (18.40)	1.20^{***} (15.43)
$PREC_{CSS}$	-0.14^{***} (-5.92)	-0.14^{***} (-5.93)	-0.11^{***} (-5.53)	-0.10^{***} (-5.99)	-0.08^{***} (-5.63)	-0.07^{***} (-5.61)	-0.10^{***} (-5.00)	-0.10^{***} (-5.01)	-0.14^{***} (-5.69)
$TMAX_{CSS}$	-1.24^{***} (-4.44)	-0.84 (-1.35)	-0.95^{***} (-3.15)	-0.09 (0.91)	-0.49^{***} (-2.98)	-0.06 (-0.74)	-0.15^{**} (-2.55)	-0.69^{***} (-3.98)	-0.80^{**} (-2.20)
$TMIN_{CSS}$	-0.47^{***} (-3.82)	-0.66^{***} (-7.05)	-0.55^{***} (-4.32)	-1.03^{***} (-9.33)	-0.86^{***} (-5.20)	-1.21^{***} (-6.96)	-0.93^{***} (-5.00)	-0.70^{***} (-3.39)	-0.76^{***} (-5.60)

Values in parentheses denote z-statistics

***, **, * $\Rightarrow$ Significance at 1%, 5%, 10% level

Table 3.14: *SUR-HEAR bootstrapped-aggregate elasticities, sorghum*

	Years								
Variable	2007	2008	2009	2010	2011	2012	2013	2014	2015
$PCORN$	−0.82*** (−5.59)	−1.22*** (−4.11)	−0.84*** (−3.57)	−0.74*** (−2.70)	−1.24*** (−4.35)	−1.21*** (−3.10)	−1.28*** (−4.93)	−1.29*** (−4.88)	−0.85*** (−2.87)
$PSOR$	1.69*** (15.08)	2.58*** (16.98)	1.57*** (17.14)	1.74*** (17.21)	2.66*** (17.09)	2.65*** (16.96)	2.51*** (17.50)	2.46*** (17.39)	1.77*** (17.06)
$PSOY$	−1.35*** (−18.47)	−2.11*** (−25.37)	−1.58*** (−24.30)	−2.21*** (−27.51)	−2.48*** (−25.31)	−2.72*** (−27.54)	−2.34*** (−25.39)	−2.20*** (−26.07)	−1.80*** (−26.64)
$PWHEAT$	0.17*** (3.92)	0.22** (2.34)	0.25** (2.39)	0.18* (1.85)	0.23*** (2.95)	0.24 (1.20)	0.30*** (2.92)	0.27*** (3.28)	0.18 (1.14)
$PDIESEL$	−0.58*** (−7.09)	−0.84*** (−5.87)	−0.57*** (−5.09)	−0.61*** (−4.20)	−0.83*** (−6.03)	−0.94*** (−4.83)	−0.99*** (−6.62)	−0.99*** (−6.26)	−0.66*** (−4.49)
$PLABOR$	0.48*** (2.93)	0.48** (2.57)	0.47*** (2.68)	0.45*** (2.71)	0.49*** (2.64)	0.44*** (2.58)	0.40** (2.00)	0.39** (2.38)	0.49*** (2.82)
AVG_{CSS}	−0.82*** (−15.62)	−0.76*** (−20.58)	−0.75*** (−22.25)	−0.88*** (−24.29)	−0.75*** (−20.25)	−0.81*** (−23.12)	−0.59*** (−17.10)	−0.56*** (−18.21)	−0.64*** (−21.33)
$PREC_{CSS}$	0.04 (1.57)	0.04* (1.93)	0.05** (2.25)	0.06** (2.31)	0.03** (2.13)	0.03** (2.09)	0.03** (2.55)	0.05** (2.41)	0.06* (1.90)
$TMAX_{CSS}$	3.54*** (14.55)	3.72*** (17.13)	3.76*** (16.38)	4.49*** (17.52)	4.15*** (16.84)	4.86*** (17.29)	4.02*** (16.18)	4.07*** (16.51)	3.97*** (17.00)
$TMIN_{CSS}$	−1.51*** (−12.72)	−1.35*** (−14.41)	−1.47*** (−13.08)	−1.75*** (−14.50)	−1.49*** (−13.76)	−1.96*** (−14.19)	−1.42*** (−13.53)	−1.50*** (−12.96)	−1.71*** (−14.19)

Values in parentheses denote z-statistics

***, **, * ⇒ Significance at 1%, 5%, 10% level

Table 3.15: *SUR-HEAR bootstrapped-aggregate elasticities, soybeans*

	Years								
Variable	2007	2008	2009	2010	2011	2012	2013	2014	2015
$PCORN$	-3.37^{***} (-16.22)	-4.55^{***} (-16.40)	-3.12^{***} (-16.37)	-2.32^{***} (-16.44)	-4.01^{***} (-16.27)	-3.92^{***} (-16.48)	-4.55^{***} (-17.08)	-4.59^{***} (-16.89)	-3.06^{***} (-16.44)
$PSOR$	0.44^{***} (2.61)	0.67^{***} (3.62)	0.43^{***} (3.02)	0.52^{***} (5.06)	0.74^{***} (3.38)	0.74^{***} (3.70)	0.76^{***} (3.74)	0.77^{***} (3.12)	0.42^{***} (2.79)
$PSOY$	4.59^{***} (40.22)	6.36^{***} (41.38)	4.52^{***} (43.83)	3.77^{***} (39.87)	5.89^{***} (42.80)	5.91^{***} (43.18)	6.78^{***} (48.09)	7.30^{***} (44.01)	5.43^{***} (43.29)
$PWHEAT$	1.09^{***} (22.94)	1.36^{***} (22.26)	1.46^{***} (22.47)	0.88^{***} (22.56)	1.22^{***} (22.24)	1.45^{***} (22.91)	1.93^{***} (22.58)	1.73^{***} (22.22)	1.25^{***} (22.60)
$PDIESEL$	-1.69^{***} (-18.49)	-2.14^{***} (-18.13)	-1.48^{***} (-18.62)	-1.32^{***} (-18.04)	-1.86^{***} (-18.82)	-2.13^{***} (-19.41)	-2.50^{***} (-18.66)	-2.61^{***} (-18.34)	-1.78^{***} (-19.17)
$PLABOR$	-0.64^{***} (-2.98)	-0.59^{***} (-3.28)	-0.64^{***} (-3.31)	-0.48^{***} (-2.97)	-0.53^{***} (-3.27)	-0.59^{***} (-3.41)	-0.79^{***} (-3.74)	-0.86^{***} (-3.41)	-0.80^{***} (-3.34)
AVG_{CSS}	3.44^{***} (35.94)	2.98^{***} (41.70)	2.77^{***} (38.50)	2.05^{***} (33.06)	2.42^{***} (35.22)	2.44^{***} (33.94)	2.56^{***} (39.31)	2.80^{***} (42.79)	2.45^{***} (36.38)
$PREC_{CSS}$	0.17^{***} (3.85)	0.14^{***} (4.48)	0.12^{***} (4.27)	0.10^{***} (4.85)	0.08^{***} (4.31)	0.07^{***} (4.41)	0.12^{***} (4.48)	0.12^{***} (4.19)	0.14^{***} (4.10)
$TMAX_{CSS}$	-6.79^{***} (-27.11)	-5.99^{***} (-26.69)	-6.18^{***} (-26.42)	-4.99^{***} (-25.84)	-5.51^{***} (-27.91)	-6.04^{***} (-26.63)	-5.85^{***} (-28.02)	-6.96^{***} (-26.10)	-6.42^{***} (-26.19)
$TMIN_{CSS}$	4.45^{***} (25.98)	3.49^{***} (24.38)	3.69^{***} (24.34)	3.30^{***} (24.61)	3.23^{***} (25.37)	3.77^{***} (26.04)	3.44^{***} (24.75)	4.14^{***} (23.78)	4.10^{***} (25.31)

Values in parentheses denote z-statistics

***, **, * $\Rightarrow$ Significance at 1%, 5%, 10% level

Table 3.16: *SUR-HEAR bootstrapped-aggregate elasticities, wheat*

	Years								
Variable	2007	2008	2009	2010	2011	2012	2013	2014	2015
$PCORN$	0.28*** (5.22)	0.47*** (6.37)	0.27*** (6.76)	0.33*** (6.70)	0.48*** (6.08)	0.47*** (6.90)	0.46*** (4.72)	0.25*** (4.83)	0.29*** (7.90)
$PSOR$	0.66 *** (11.21)	1.04 *** (11.11)	0.65 *** (10.54)	0.86 *** (12.00)	1.15 *** (11.15)	1.20 *** (11.03)	1.12 *** (10.99)	1.10 *** (11.62)	0.72*** (9.84)
$PSOY$	−1.19*** (−18.44)	−2.00*** (−23.51)	−1.40*** (−20.11)	−2.12*** (−23.82)	−2.40*** (−23.90)	−2.60*** (−24.00)	−2.42*** (−23.06)	−2.07*** (−23.58)	−1.65*** (−22.61)
$PWHEAT$	−0.05*** (−4.10)	−0.10*** (−5.99)	−0.08*** (−5.28)	−0.07*** (−5.41)	−0.10*** (−5.02)	−0.15*** (−6.83)	−0.19*** (−4.91)	−0.09*** (−3.86)	−0.10*** (−7.10)
$PDIESEL$	0.19*** (5.88)	0.28*** (7.08)	0.18*** (7.11)	0.25*** (7.60)	0.29*** (6.68)	0.31*** (7.16)	0.29*** (5.64)	0.18*** (5.75)	0.21*** (8.43)
$PLABOR$	−0.23** (−2.31)	−0.24** (−2.27)	−0.30* (−1.81)	−0.30 (−0.86)	−0.27** (−2.21)	−0.34* (−1.93)	−0.38*** (−3.22)	−0.43** (−2.53)	−0.37 (−1.54)
AVG_{CSS}	−0.69*** (−15.92)	−0.83*** (−19.14)	−0.76*** (−18.05)	−1.04*** (−22.95)	−0.88*** (−20.13)	−0.80*** (−20.56)	−0.79*** (−16.89)	−0.58*** (−16.03)	−0.67*** (−18.43)
$PREC_{CSS}$	0.02*** (5.10)	0.02*** (3.63)	0.04*** (4.23)	0.03*** (4.23)	0.01*** (3.74)	0.02*** (4.69)	0.01*** (4.74)	0.03*** (5.08)	0.02*** (3.37)
$TMAX_{CSS}$	1.30 *** (11.36)	1.65 *** (14.56)	1.41 *** (12.85)	1.72 *** (15.19)	1.87 *** (14.37)	2.07 *** (15.41)	1.98 *** (13.45)	1.78 *** (13.65)	1.71 *** (14.18)
$TMIN_{CSS}$	−0.28*** (−6.35)	−0.46*** (−8.51)	−0.32*** (−6.85)	−0.42*** (−9.31)	−0.46*** (−7.37)	−0.48*** (−9.42)	−0.52*** (−7.55)	−0.45*** (−6.44)	−0.46*** (−8.40)

Values in parentheses denote z-statistics

***, **, * $\Rightarrow$ Significance at 1%, 5%, 10% level

Table 3.17: *ANN bagged-aggregate elasticities, corn*

	Years							
Variable	2008	2009	2010	2011	2012	2013	2014	2015
$CORN_{t-1}$	0.02^{***} (7.75)	0.01^{*} (1.89)	0.01^{**} (2.25)	0.02^{***} (4.89)	0.02^{***} (6.83)	0.02^{***} (5.47)	0.00 (1.11)	0.00 (1.55)
SOR_{t-1}	0.00^{**} (2.53)	0.00 (1.31)	0.00 (1.09)	0.01^{***} (4.19)	0.00^{**} (2.00)	0.01^{***} (3.32)	0.01^{**} (2.06)	0.01^{**} (2.00)
SOY_{t-1}	-0.01^{*} (-1.86)	-0.01 (-1.15)	-0.01 (-1.40)	-0.01^{**} (-2.57)	-0.01^{***} (-2.87)	-0.01^{*} (-1.84)	-0.01^{***} (-2.76)	0.00 (-1.02)
WHT_{t-1}	-0.02^{***} (-10.57)	-0.01^{***} (-9.80)	-0.01^{***} (-8.52)	-0.01^{***} (-7.95)	-0.01^{***} (-8.66)	-0.01^{***} (-6.38)	-0.01^{***} (-5.65)	-0.01^{***} (-4.44)
$PCORN$	0.10^{**} (1.97)	0.03 (0.76)	0.04 (1.36)	0.08^{*} (1.82)	0.05 (1.41)	0.09^{*} (1.87)	0.05 (1.62)	0.04 (1.18)
$PSOR$	-0.06 (-1.23)	-0.04^{*} (-1.83)	-0.04 (-1.55)	-0.03 (-0.90)	-0.04 (-1.23)	0.01 (0.33)	-0.05^{*} (-1.75)	-0.02 (-0.75)
$PSOY$	0.00 (0.06)	-0.03 (-0.94)	0.01 (0.40)	0.00 (-0.02)	0.02 (0.39)	0.05 (1.30)	-0.01 (-0.17)	-0.03 (-0.93)
$PWHEAT$	-0.02 (-0.50)	-0.01 (-0.28)	-0.02 (-0.71)	-0.03 (-0.83)	-0.04 (-1.30)	-0.13^{**} (-2.33)	-0.05 (-1.12)	-0.03 (-0.60)
$PDIESEL$	0.04 (0.29)	0.03 (0.35)	0.03 (0.34)	0.03 (0.29)	0.02 (0.13)	-0.17 (-1.04)	-0.02 (-0.22)	0.05 (0.51)
$PLABOR$	-0.03 (-0.21)	-0.15 (-0.57)	-0.10 (-0.45)	0.06 (0.46)	-0.08 (-0.34)	-0.11 (-0.36)	-0.12 (-0.44)	-0.06 (-0.21)
AVG_{CSS}	0.01 (1.32)	0.00 (-0.28)	0.00 (-0.07)	0.00 (0.40)	0.01 (1.06)	0.02^{*} (1.94)	0.00 (-0.33)	0.00 (-0.13)
AVG_{W}	-0.02 (-1.51)	-0.04^{***} (-3.78)	-0.02 (-1.37)	-0.01 (-0.43)	-0.01 (-1.07)	-0.06^{***} (-3.47)	-0.01 (-1.35)	-0.02 (-1.24)
$PREC_{CSS}$	0.01^{***} (2.65)	0.01^{*} (1.69)	0.01 (1.17)	0.01^{***} (2.86)	0.01^{*} (1.77)	0.00 (0.70)	0.01 (1.63)	0.01 (0.75)
$PREC_{W}$	0.00 (-0.45)	0.01^{*} (1.95)	0.01 (1.14)	0.00 (-0.30)	0.00 (-0.61)	0.00 (-0.95)	0.00 (-0.45)	0.00 (-0.25)
$TMAX_{CSS}$	0.04 (0.57)	-0.03 (-0.36)	0.00 (0.02)	0.08 (1.16)	0.06 (0.68)	0.12 (1.24)	0.05 (0.63)	0.01 (0.06)
$TMAX_{W}$	-0.05 (-0.52)	-0.08 (-0.87)	-0.05 (-0.89)	-0.08 (-1.04)	-0.13^{*} (-1.82)	-0.10 (-0.98)	0.04 (0.51)	-0.05 (-0.62)
$TMIN_{CSS}$	-0.01 (-0.43)	-0.02 (-0.31)	-0.04 (-0.83)	-0.04 (-1.43)	-0.07^{*} (-1.89)	-0.04 (-0.87)	0.00 (-0.04)	-0.06 (-1.24)
$TMIN_{W}$	-0.01 (-0.46)	-0.01 (-0.14)	0.02 (0.54)	0.01 (0.26)	0.02 (0.84)	0.00 (0.18)	0.00 (-0.06)	0.02 (0.54)

Values in parentheses denote z-statistics

***, **, * $\Rightarrow$ Significance at 1%, 5%, 10% level

Table 3.18: *ANN bagged-aggregate elasticities, sorghum*

	Years							
Variable	2008	2009	2010	2011	2012	2013	2014	2015
$CORN_{t-1}$	0.00* −1.73	0.00 0.15	0.00 0.74	0.00 −1.37	0.00 −1.55	0.00 −0.87	0.00 −0.25	0.00 −0.60
SOR_{t-1}	0.00 −0.36	−0.01*** −3.86	−0.01*** −6.30	−0.02*** −12.64	−0.02*** −9.97	−0.02*** −9.58	−0.01*** −13.38	−0.01*** −10.46
SOY_{t-1}	−0.01 −1.14	0.00 0.55	0.00 −0.18	0.00 −1.09	0.00 −0.60	−0.01 −0.80	0.00 −0.13	0.00 −0.55
WHT_{t-1}	−0.02*** −4.26	−0.02*** −4.30	−0.01*** −3.74	−0.01*** −4.29	−0.01*** −5.70	−0.01*** −4.28	−0.01*** −4.97	−0.01*** −4.46
$PCORN$	0.05 0.68	0.01 0.18	−0.01 −0.15	0.07 0.99	0.04 0.73	0.11 1.64	0.05 1.11	0.03 0.61
$PSOR$	0.09 1.08	0.00* −0.12	0.02 0.53	0.08 1.15	0.04 0.76	0.07 1.16	0.03 0.55	0.02 0.36
$PSOY$	−0.05 −0.71	−0.04 −0.92	−0.06 −1.24	−0.03 −0.37	−0.03 −0.52	−0.06 −1.62	−0.03 −0.59	−0.01 −0.31
$PWHEAT$	0.03 0.39	−0.01 −0.10	0.01 0.18	0.06 0.82	0.06 1.10	0.08 1.10	0.05 0.77	0.01 0.21
$PDIESEL$	0.03 0.17	0.00 −0.03	−0.03 −0.19	0.04 0.22	0.09 0.40	−0.06 −0.31	−0.04 −0.22	−0.02 −0.14
$PLABOR$	0.25 1.07	0.31 0.73	0.20 0.53	0.21 1.08	−0.13 −0.37	−0.01 −0.02	−0.26 −0.61	0.09 0.15
AVG_{CSS}	−0.02 −1.34	−0.03 −1.56	−0.03* −1.91	−0.03** −2.07	−0.02 −0.94	0.00 0.29	−0.02* −1.71	−0.03 −1.63
AVG_{W}	0.00 −0.24	0.02 1.16	0.01 0.41	0.01 0.88	−0.02 −1.02	0.00 0.05	0.01 0.78	0.02 1.17
$PREC_{CSS}$	0.01 1.26	0.01* 0.67	0.01 0.93	0.01 2.13	0.01 1.49	0.00 0.58	0.01* 1.58	0.01 0.83
$PREC_{W}$	0.00 −0.76	0.01 0.93	0.00 0.50	0.00 −0.49	0.00 −0.69	−0.01 −1.32	−0.01 −1.08	0.00 −0.65
$TMAX_{CSS}$	−0.04 −0.42	0.05 0.41	0.08 0.62	−0.11 −1.03	−0.23 −1.41	0.08 0.64	−0.08 −0.66	−0.17 −0.97
$TMAX_{W}$	0.03 0.24	−0.08 −0.72	−0.03 −0.42	0.01 0.04	−0.01 −0.08	0.03 0.19	0.16 1.15	−0.05 −0.36
$TMIN_{CSS}$	−0.05 −0.97	−0.08 −1.35	−0.06 −1.00	−0.05 −0.85	0.00 −0.05	−0.07 −1.48	0.01 0.17	−0.05 −0.79
$TMIN_{W}$	0.01 0.14	0.04 0.86	0.04 0.97	0.01 0.12	0.03 0.53	−0.01 −0.46	−0.01 −0.16	0.03 0.54

Values in parentheses denote z-statistics

***, **, * $\Rightarrow$ Significance at 1%, 5%, 10% level

Table 3.19: *ANN bagged-aggregate elasticities, soybeans*

	Years							
Variable	2008	2009	2010	2011	2012	2013	2014	2015
$CORN_{t-1}$	-0.01^{***} (-2.93)	-0.01^{**} (-2.29)	-0.01^{***} (-2.70)	-0.01^{***} (-2.98)	-0.02^{***} (-4.07)	0.00 (0.13)	-0.01^{***} (-3.86)	-0.01^{**} (-2.29)
SOR_{t-1}	0.00 (-0.84)	0.00 (0.67)	0.00 (0.07)	0.00 (0.81)	0.00 (0.98)	0.00 (1.17)	0.00 (0.52)	0.00 (0.72)
SOY_{t-1}	0.02^{***} (3.94)	0.01^{**} (2.07)	0.02^{***} (2.65)	0.03^{***} (4.12)	0.03^{***} (5.21)	0.01 (0.99)	0.00^{*} (-1.69)	0.01^{***} (3.21)
WHT_{t-1}	-0.03^{***} (-7.90)	-0.02^{***} (-8.46)	-0.02^{***} (-8.64)	-0.02^{***} (-8.29)	-0.02^{***} (-9.60)	-0.01^{***} (-6.45)	-0.02^{***} (-8.59)	-0.02^{***} (-6.24)
$PCORN$	0.04 (0.58)	0.01 (0.23)	0.01 (0.25)	-0.02 (-0.50)	-0.04 (-0.81)	-0.06 (-0.90)	-0.06 (-1.33)	0.01 (0.16)
$PSOR$	-0.05 (-1.04)	0.01 (0.46)	0.00 (-0.06)	-0.01 (-0.20)	0.02 (0.55)	0.10^{*} (1.92)	-0.01 (-0.18)	0.02 (0.71)
$PSOY$	0.02 (0.32)	0.02 (0.69)	0.03 (1.03)	0.01 (0.18)	0.03 (0.53)	0.08 (1.39)	0.01 (0.29)	0.03 (1.00)
$PWHEAT$	-0.04 (-0.98)	-0.01 (-0.29)	0.03 (1.00)	-0.03 (-0.74)	0.00 (-0.03)	-0.31^{***} (-3.79)	-0.06 (-1.27)	0.01 (0.29)
$PDIESEL$	-0.05 (-0.43)	0.01 (0.13)	-0.01 (-0.20)	-0.04 (-0.46)	-0.12 (-0.94)	-0.38 (-1.54)	-0.08 (-0.65)	0.07 (0.78)
$PLABOR$	-0.13 (-0.53)	-0.14 (-0.52)	-0.05 (-0.19)	0.08 (0.55)	0.31 (1.25)	0.10 (0.27)	0.22 (0.74)	-0.02 (-0.08)
AVG_{CSS}	0.01 (0.99)	0.02 (1.50)	0.01 (0.86)	0.00 (0.08)	0.02 (1.32)	0.02 (1.10)	0.01 (0.53)	-0.01 (-0.43)
AVG_{W}	0.00 (0.13)	0.02 (1.27)	0.04^{***} (2.67)	0.01 (0.79)	0.03^{*} (1.73)	-0.07^{***} (-2.84)	0.00 (-0.01)	0.04^{**} (2.22)
$PREC_{CSS}$	0.01 (0.96)	0.00 (-0.53)	0.00 (0.44)	0.00 (-0.09)	0.00 (-0.79)	-0.01 (-1.45)	-0.01 (-1.02)	-0.01 (-0.84)
$PREC_{W}$	0.01 (1.07)	0.01 (0.62)	0.00 (0.74)	0.00 (-0.25)	0.00 (-0.36)	0.00 (0.70)	0.00 (0.31)	-0.01 (-0.91)
$TMAX_{CSS}$	-0.13 (-1.36)	-0.17^{*} (-1.82)	-0.15^{*} (-1.83)	-0.02 (-0.28)	-0.02 (-0.22)	0.28^{**} (1.99)	-0.02 (-0.25)	-0.05 (-0.44)
$TMAX_{W}$	-0.19^{*} (-1.87)	-0.21^{***} (-2.63)	-0.10 (-1.52)	-0.21^{***} (-2.79)	-0.20^{**} (-2.55)	-0.43^{***} (-3.53)	-0.16^{*} (-1.81)	-0.21^{**} (-2.32)
$TMIN_{CSS}$	0.06 (1.19)	0.08 (1.55)	0.04 (0.89)	0.04 (0.97)	0.03 (0.54)	0.05 (0.75)	0.09^{**} (2.00)	0.06 (0.92)
$TMIN_{W}$	0.00 (-0.01)	0.00 (0.11)	0.01 (0.24)	0.00 (0.05)	-0.02 (-0.76)	-0.02 (-0.45)	0.00 (-0.07)	0.01 (0.12)

Values in parentheses denote z-statistics

***, **, * $\Rightarrow$ Significance at 1%, 5%, 10% level

Table 3.20: *ANN bagged-aggregate elasticities, wheat*

	Years							
Variable	2008	2009	2010	2011	2012	2013	2014	2015
$CORN_{t-1}$	0.00 *** (−3.28)	0.00 (−0.10)	0.00 (0.50)	0.00 *** (−3.14)	0.00 *** (−3.27)	0.00 ** (−2.48)	0.00 (0.58)	0.00 (0.32)
SOR_{t-1}	0.00 (−1.12)	0.00 (0.55)	0.00 (0.67)	0.00 (1.13)	0.00 (1.28)	0.00 (1.58)	0.00 (1.13)	0.00** (2.12)
SOY_{t-1}	0.00 (0.32)	0.00 (0.01)	0.00 (−1.19)	0.00 (−0.56)	0.00 (−0.07)	0.00 (−0.10)	0.00 (1.25)	0.00 (0.02)
WHT_{t-1}	0.03 *** (12.13)	0.02 *** (13.97)	0.02 *** (13.40)	0.02 *** (13.36)	0.02 *** (16.41)	0.02 *** (13.28)	0.02 *** (13.02)	0.02 *** (11.73)
$PCORN$	−0.02 (−0.59)	0.00 (0.15)	−0.01 (−0.59)	−0.01 (−0.41)	0.00 (−0.20)	−0.02 (−0.81)	−0.01 (−0.57)	−0.01 (−0.36)
$PSOR$	0.00 (−0.16)	0.00 (0.17)	0.01 (0.70)	−0.01 (−0.24)	−0.02 (−0.73)	−0.01 (−0.46)	0.00 (−0.22)	−0.01 (−0.33)
$PSOY$	0.00 (−0.07)	0.02 (0.97)	0.01 (0.53)	0.00 (0.01)	−0.01 (−0.32)	0.01 (0.63)	0.01 (0.45)	0.02 (0.96)
$PWHEAT$	0.01 (0.63)	0.03 (0.99)	0.01 (0.47)	0.02 (0.85)	0.02 (0.97)	0.01 (0.34)	0.03 (0.95)	0.02 (0.75)
$PDIESEL$	−0.05 (−0.61)	0.01 (0.21)	−0.01 (−0.08)	−0.06 (−0.79)	−0.04 (−0.49)	−0.06 (−0.62)	−0.05 (−0.59)	−0.03 (−0.45)
$PLABOR$	0.05 (0.50)	0.02 (0.10)	0.05 (0.34)	0.01 (0.16)	−0.05 (−0.42)	−0.03 (−0.19)	−0.07 (−0.53)	−0.10 (−0.50)
AVG_{CSS}	0.01 (0.92)	0.01 (1.14)	0.01 (1.32)	0.01 (1.34)	0.00 (0.33)	0.00 (0.05)	0.01 (1.15)	0.01 (0.94)
AVG_{W}	0.00 (0.28)	−0.01 (−1.13)	−0.01 (−1.28)	0.00 (−0.13)	0.00 (−0.42)	−0.01 (−0.70)	0.00 (−0.61)	0.00 (−0.35)
$PREC_{CSS}$	0.00 (−1.34)	0.00 (−0.63)	0.00 (−0.75)	0.00 (−1.39)	0.00 (−1.04)	0.00 (−0.94)	0.00 (−1.01)	−0.01 (−1.07)
$PREC_{W}$	0.00 (0.45)	−0.01 (−1.39)	0.00 (−1.32)	0.00 (1.04)	0.00 (0.91)	0.00 (0.97)	0.00 (0.84)	0.00 (0.07)
$TMAX_{CSS}$	0.02 (0.37)	0.08 (1.43)	0.04 (0.78)	0.01 (0.22)	0.03 (0.55)	0.00 (−0.01)	−0.01 (−0.19)	0.04 (0.48)
$TMAX_{W}$	0.01 (0.18)	0.06 (1.18)	0.01 (0.33)	0.04 (0.73)	0.05 (1.06)	0.03 (0.50)	0.01 (0.17)	0.03 (0.52)
$TMIN_{CSS}$	−0.01 (−0.29)	−0.01 (−0.21)	0.00 (0.09)	0.00 (0.18)	0.02 (0.75)	−0.02 (−0.72)	−0.01 (−0.27)	0.01 (0.30)
$TMIN_{W}$	0.02 (0.92)	0.01 (0.30)	0.00 (−0.04)	0.02 (0.83)	0.01 (0.84)	0.01 (1.03)	0.02 (0.78)	0.01 (0.28)

Values in parentheses denote z-statistics

***, **, * $\Rightarrow$ Significance at 1%, 5%, 10% level

Chapter 4

Neural network estimators of binary choice processes for valuing environmental amenities: Estimation, marginal effects, and WTP

4.1 Introduction

Discrete choice analysis involves the modeling of a behavioral process whereby an agent makes or selects a choice or option from a discrete set of alternatives. Estimated discrete choice econometric models try to represent the behavioral process conditional on a number of explanatory factors in order to estimate the probability of making or picking a particular choice or option. In its simplest form, the dependent variable of such a model is binary (e.g., Yes/No). While predicting the probability of an individual selecting a particular choice is of interest, researchers are also interested in more substantive inquiries offered by discrete choice analysis. For example, the goal of a study may be to explore not only individuals' probabilities of making alternative choices, but rather how and what factors impact these probabilities within the sample population. Such substantive inference is usually examined using marginal effects (Train, 2003).

In a contingent-valuation framework (CV), discrete choice analysis is often used to estimate the probability of an individual voting in favor of a proposed policy or taking a particular environmental related action regarding a non-market good. Included within the set of explanatory variables is typically a payment vehicle, through which the cost of the action is imposed upon the individual or the individual receives some form of payment. Through the use of binary choice models, CV studies can provide a measure of an individual's willingness-to-pay (WTP) or accept (WTA) to protect, enhance or conserve a given environmental amenity or resource via a proposed policy or choice. In these studies the WTP (WTA) measure and related inferences are often of greater importance than the predicted probabilities (Hanemann, 1984). For the purpose of this study, we focus on WTP, though WTA can be obtained with modest changes in the methods presented[1].

Estimation of binary choice models typically requires that the econometric model satisfy the utility maximization hypothesis. The most widely used models for this purpose are the binary logit and probit models. To satisfy the utility maximization hypothesis, the argument (or index function) of these models must be able to be interpreted as the difference in utility between two states of existence defined by the dependent variable. This requirement provides a practical procedure for specifying the functional form of the index function by postulating a priori the underlying functional form of a representative utility function (Hanemann, 1984). However, an a priori imposition of a theoretical structure on a statistical model without considering the underlying probabilistic structure of the observed data can leave the estimable model statistically misspecified.

One method to avoid potential misspecification is to weaken the distributional assumptions of the models and rely upon semi-nonparametric (SNP) techniques. Cooper (2002) provides an overview of SNP approaches applied to dichotomous-choice models. Examples

[1] For example, Horowitz (1993) and Sugden (1999) show that, in theory, $\frac{\partial WTP}{\partial y} = 1 - \frac{WTP}{WTA}$ where y is the individual's income.

include Gallant and Nychka (1987) who use a flexible distribution-based approach using Hermite polynomial expansions or Creel and Loomis (1997) who use a Fourier functional form for the index (predictor) function of the model. Another SNP approach that has been applied to dichotomous-choice models is that of Klein and Spady (1993). The Klein and Spady (1993) estimator makes no assumptions regarding the distribution of disturbances but does depend on a parametrically-specified index function (Klein & Spady, 1993). A less well known SNP technique is an artificial neural network (ANN). ANNs have been used for classification problems and have the ability to learn arbitrary and highly nonlinear functional mappings using finite data (Mehrotra, Mohan, & Ranka, 1997). Hornik, Stinchcombe, and White (1989) show that feed-forward back-propagation artificial neural networks (FFBANNs) can act as universal function approximators under fairly general conditions. Ripley (1994) further concludes that this result is easily extended to networks that are used to model binary-choice processes.

The purpose of this paper is to examine the estimation of binary-choice processes using ANNs in a CV context. In addition, given the "black box" nature of the interpretation of the estimable parameters of an ANN, this paper expands the literature by deriving and estimating marginal effects of explanatory variables on the probability of making a choice using ANNs. To the authors' knowledge, this is a novel contribution to the literature on using ANNs in a regression type framework. We also provide a simple algorithm that can be used to estimate median WTP in CV (type) studies, providing a method for estimating consistent WTP measures using FFBANNs. The results are compared to estimates from logit and probit models and — due to its availability in existing software — the Klein and Spady (1993) estimator. Comparisons are made using CV survey data from two studies, which allows for comparison of WTP estimates, as well. The first study is based on survey data collected by a research team (that included the author) in the Smoky Hill Watershed region of Kansas. This is a new study that examines community members' WTP via increased water bills to maintain water usage during times of drought. The second study uses data made publicly available by Calderon, Anit, Palao, and Lasco (2012) and examines WTP via higher water bills to fund conservation projects in the Layawan Watershed in the Philippines. We believe the methods and results from this paper will not only help to advance modeling of CV survey data, but will be highly applicable for stated-choice survey data and other discrete choice modeling problems. Thus, the paper has broader implications than just the CV analysis and examples presented.

The remainder of the paper is structured as follows. Section 4.2 provides background on CV of non-market goods and SNP estimation of binary discrete choice models for CV. Section 4.3 introduces the FFBANN regression function as a SNP flexible functional form and provides an overview of network estimation, as well as the estimation of marginal effects and WTP when using FFBANNs. In Section 4.4, two empirical applications of FFBANNs are presented and compared to simple logit and probit analysis and the Klein and Spady (1993) approach. Results are presented in Section 4.5 and Section 4.5 offers some concluding remarks.

4.2 Contingent-Valuation Modeling and Semi-nonparametric Methods

Hanemann (1984) provides the basis for modeling CV survey data with binary responses. Following Hanemann (1991), consider an individual who derives utility from the supply of some environmental amenity. Let q denote the supply of the amenity; I the individual's income; and $\mathbf{s}$ a vector of variables representing the consumption of other market commodities, prices, demographic characteristics and other attributes of the individual. The individual's indirect-utility function is then given by $V_a(q_a, I, \mathbf{s})$ where V_a is the observable component of the indirect-utility function, and a is an index denoting the amount of q being consumed.

Consider the situation where the individual is faced with the opportunity of increasing consumption of q from q_0 to q_1. If the increase in q costs C, the individual will pay the amount if:

$$V_1\left(q_1, I-C, \mathbf{s}\right) \geq V_1\left(q_0, I, \mathbf{s}\right). \tag{4.1}$$

The individual's maximum WTP (C_p) — equal to the compensating-variation measure of the change in q — is found where $V_1\left(q_1, I-C, \mathbf{s}\right) = V_0\left(q_0, I, \mathbf{s}\right)$ (Hanemann, 1991).

In practice, the individual's decision to pay \$$C$ is observable but his utility contains unobservable components and is treated as stochastic (Hanemann, 1984). Thus, the individual's indirect utility is decomposed as:

$$V_a\left(q_a, I, \mathbf{s}, \varepsilon_a\right) = v_a\left(q_a, I, \mathbf{s}\right) + \varepsilon_a \tag{4.2}$$

where ε_a is an *IID* random variable with zero mean (An, 2000; Hanemann, 1984). From this perspective, the individual's response can be viewed in a probabilistic framework:

$$p = P\left[\text{individual pays } \$C \text{ to increase } q\right] \tag{4.3}$$

$$p = P\left[V_1\left(q_1, I-C, \mathbf{s}, \varepsilon_1\right) \geq V_0\left(q_0, I, \mathbf{s}, \varepsilon_0\right)\right] \tag{4.4}$$

$$p = P\left[v_1\left(q_1, I-C, \mathbf{s}\right) + \varepsilon_1 \geq v_0\left(q_0, I, \mathbf{s}\right) + \varepsilon_0\right] \tag{4.5}$$

$$p = P\left[\Delta v \geq \eta\right] \tag{4.6}$$

where p represents the probability that the offer is accepted, $\Delta v = v_1(\cdot) - v_0(\cdot)$, and $\eta = \varepsilon_0 - \varepsilon_1$. Based on this, equation 4.6 can be written as:

$$p = F_\eta\left(\Delta v\right), \tag{4.7}$$

where $F_\eta(\cdot)$ is the cumulative distribution function (cdf) of η (Hanemann, 1984). Thus, as stated by Hanemann (1984, p. 334), "if the statistical binary response model is to be interpreted as the outcome of a utility-maximizing choice, the argument of $F_\eta(\cdot)$... must take the form of a utility difference [i.e., Δv]." This approach provides a mechanism to determine if a given statistical model is compatible with the utility maximization procedure for specifying a theoretically consistent functional form for a given model (Hanemann, 1984). Once Δv has been specified, the modeler need only specify $F_\eta(\cdot)$, which is dependent upon the assumed distributions of ε_0 and ε_1[2].

A weakness of this approach is that the researcher has to make an assumption about the distribution of the stochastic term, which is usually unknown (Cosslett, 1983). Because the researcher only observes the response by the individual to the offer of $\$C$ to increase q, the response should be empirically viewed as a Bernoulli random variable with parameter p, which represents the probability of a response of yes or accept (Powers & Xie, 2008). Let y_i denote the response by the i^{th} individual, where

$$y_i = \begin{cases} 1 & \text{for ``yes'' or ``accept''} \\ 0 & \text{otherwise} \end{cases}. \tag{4.8}$$

Assume that y_i is dependent upon a $m \times 1$ vector of unknown explanatory factors, $\mathbf{x}_i$, via the following relationship:

$$E\left(y_i \mid \mathbf{X} = \mathbf{x}_i\right) = F_\eta\left[\mathcal{I}\left(\mathbf{x}_i; \boldsymbol{\beta}\right)\right], \tag{4.9}$$

where $F_\eta(\cdot) : R \rightarrow [0,1]$ (a transformation function), $\mathcal{I}(\cdot) : R^m \rightarrow R$ (a predictor or index function), and $\boldsymbol{\beta}$ is a $m \times 1$ vector of unknown parameters (Amemiya, 1981; Davidson, MacKinnon, et al., 1993). Common choices for $F_\eta(\cdot)$ are the logistic and standard normal cumulative distribution functions.

Bergtold et al. (2010) show that equation 4.9 will give rise to a proper regression model if the conditional-Bernoulli distribution underlying the conditional mean given by equation 4.9 can be derived from a proper joint-density function of y_i and $\mathbf{x}_i$. Arnold and Press (1989)

[2]When ε_0 and ε_1 are *IID* extreme value, then $F_\eta(\cdot)$ is the logistic cdf. When ε_0 and ε_1 are *IID* normal, $F_\eta(\cdot)$ is the normal cdf (Train, 2003).

show that the existence of the logistic-regression model (i.e., existence of a proper joint-density function) in observational situations depends on the compatibility between the conditional distribution $f(y_i \mid \mathbf{x}_i; \boldsymbol{\beta})$ and the inverse-conditional distribution $f(\mathbf{x}_i \mid y_i; \boldsymbol{\theta})$. That is, $f(y_i \mid \mathbf{x}_i; \boldsymbol{\beta}) f_{\mathbf{X}}(y_i, \mathbf{x}_i; \boldsymbol{\vartheta}) = f(\mathbf{x}_i \mid y_i; \boldsymbol{\theta}) f_Y(y_i, \mathbf{x}_i; p) = f(y_i, \mathbf{x}_i; \boldsymbol{\varphi})$ where $f_{\mathbf{X}}(y_i, \mathbf{x}_i; \boldsymbol{\vartheta})$ is the multivariate-marginal distribution of $\mathbf{X}$, $f_Y(y_i, \mathbf{x}_i; p)$ is the marginal distribution of y_i, $f(y_i, \mathbf{x}_i; \boldsymbol{\varphi})$ is the multivariate distribution of Y and $\mathbf{X}$, and $\boldsymbol{\vartheta}$ and $\boldsymbol{\varphi}$ are appropriate sets of parameters. Bergtold et al. (2010) show that the logistic-regression formulation arises naturally from this model specification approach, giving a predictor function of the form (Bergtold et al., 2010, p. 7):

$$\mathcal{I}(\mathbf{x}_i; \boldsymbol{\beta}) = \ln\left[\frac{f(\mathbf{x}_i \mid y_i = 1; \boldsymbol{\theta})}{f(\mathbf{x}_i \mid y_i = 0; \boldsymbol{\theta})}\right] + \ln\left[\frac{P(y_i = 1)}{P(y_i = 0)}\right]. \tag{4.10}$$

Kay and Little (1987) show that if

$$\ln\left[\frac{f(\mathbf{x}_i \mid y_i = 1; \boldsymbol{\theta})}{f(\mathbf{x}_i \mid y_i = 0; \boldsymbol{\theta})}\right] = \alpha_0 + \boldsymbol{\alpha}'\mathbf{g}(\mathbf{x}_i), \tag{4.11}$$

then

$$\ln\left[\frac{f(y_i = 1 \mid \mathbf{x}_i; \boldsymbol{\beta})}{f(y_i = 0 \mid \mathbf{x}_i; \boldsymbol{\beta})}\right] = \gamma_0 + \boldsymbol{\gamma}'\mathbf{g}(\mathbf{x}_i) \tag{4.12}$$

where $\gamma_0 = \alpha_0 + \ln\left[\frac{P(y_i=1)}{P(y_i=0)}\right]$, $\boldsymbol{\gamma} = \boldsymbol{\alpha}$, and $\mathbf{g}(\mathbf{x}_i)$ is a vector of suitable transformations of $\mathbf{x}_i$, the vector of explanatory variables. This gives rise to the following functional form for the logistic-regression model:

$$E(y_i \mid \mathbf{X} = \mathbf{x}_i) = [1 + \exp(-\gamma_0 - \boldsymbol{\gamma}\mathbf{g}(\mathbf{x}_i))]^{-1}. \tag{4.13}$$

Kay and Little (1987) provide the transformation functions $\mathbf{g}(\mathbf{x}_i)$ required to satisfy the logistic model when the conditional distribution of $\mathbf{X}$ is a member of the simple exponential family.

The work by Kay and Little (1987) emphasizes the dependency between the functional forms of the index and transformation functions. For example, when $F(\mathbf{x}_i \mid y_i = j; \boldsymbol{\theta})$ for $j = 0, 1$, is multivariate normal with heterogeneous covariance matrix dependent upon j and $F_\eta(\cdot)$ is the logistic cdf, the index function $\mathcal{I}(\mathbf{x}_i; \boldsymbol{\beta})$ is a quadratic function of $\mathbf{x}_i$. If the covariance matrix is homogeneous, the index function is linear in $\mathbf{x}_i$ (Kay & Little, 1987). Kay and Little (1987, p. 498) state that "in cases other than multivariate normality, however, little can be said since there are few other multivariate distributions which could act as appropriate models" and that provide linear-index functions. In light of this, Arnold and Press (1989) question many of the binary-choice models presented in the literature.

The choice of which functional form to use for the index and transformation functions concerns the parameterization of the contemporaneous dependence between y_i and $\mathbf{x}_i$ (Spanos, 1999). Given that researchers have the ability to vary the functional form of $\mathcal{I}(\mathbf{x}_i; \boldsymbol{\beta})$, Amemiya (1981) states that the importance of having $F_\eta(\cdot)$ correctly specified is lessened. If one can approximate $\mathcal{I}(\mathbf{x}_i; \boldsymbol{\beta})$ for a given choice of $F_\eta(\cdot)$, then the particular choice of $F_\eta(\cdot)$ need only satisfy the conditions of a transformation function. As compelling as this argument is, a particular choice of $F_\eta(\cdot)$ may not give rise to a proper statistical model in the sense that the conditional-Bernoulli distribution based upon $F_\eta(\cdot)$ cannot be derived from a proper joint-density function. A choice of $F_\eta(\cdot)$ that does allow for the approximation of $\mathcal{I}(\mathbf{x}_i; \boldsymbol{\beta})$ is the logistic cdf (Bergtold et al., 2010). Thus, one way of weakening the functional-form (and also distributional) assumptions is to employ semi-nonparametric (SNP) estimation methods within the logistic-regression framework.

SNP methods are semi-distribution free approaches that avoid restricting $F_\eta(\cdot)$ and/or $\mathcal{I}(\mathbf{x}_i; \boldsymbol{\beta})$ in equation 4.10 by trying to estimate the compound function $F_\eta[\mathcal{I}(\mathbf{x}_i; \boldsymbol{\beta})]$ (Cooper, 2002). Following Cooper (2002), the modeler can replace $F_\eta(\cdot)$, $\mathcal{I}(\mathbf{x}_i; \boldsymbol{\beta})$, or both with a flexible SNP functional form. Results from Gabler, Laisney, and Lechner (1993) and Horowitz (1993) suggest that SNP estimation may help in avoiding model misspecification due to an incorrect functional form. A SNP approach may be advantageous if the resulting predictor function is not easily specifiable as indicated by Kay and Little (1987) or it is highly nonlinear.

A SNP estimator that can be found throughout the dichotomous-choice literature is that of Creel and Loomis (1997), which estimates the compound function $F_\eta\left[\mathcal{I}\left(\mathbf{x}_i;\boldsymbol{\beta}\right)\right]$ using a flexible-Fourier functional form. This estimator has been used to value the reduction of risk exposure to hazardous waste (Creel & Loomis, 1997), to estimate farmer premiums for conservation adoption (Cooper & Signorello, 2008), and was extended to a multivariate-discrete choice by Cooper (2003) to examine farmers' willingness to adopt a bundle of conservation practices. Hermite-polynomial approaches similar to that of Gallant and Nychka (1987) have been used, for example, to estimate WTP for water supply improvements (Arouna & Dabbert, 2012) and the willingness of producers to use eco-labels (Chang, 2012). The distribution-free estimator of Klein and Spady (1993) has been used to estimate WTP for sanitation improvements (Adriano, Wilson, & Joao, 2011) and the valuation of time (Bastin, Cirillo, & Toint, 2010; Fosgerau, 2006; Fosgerau et al., 2005). One SNP approach, however, that is yet to be widely applied in the field of natural resource and environmental economics, and more specifically in the analysis of CV or stated-choice survey data, is the feed-forward back-propagation artificial neural network, which provides a potentially powerful SNP tool for modeling dichotomous-choice CV models. Furthermore, this approach can be extended to many other binary-discrete-choice modeling frameworks.

4.3 Feed-Forward Back-Propagation Artificial Neural Networks

4.3.1 Functional Specification (Network Architecture)

Fausett (1994, p. 3) defines an artificial neural network (ANN) as "an information-processing system that has certain performance characteristics in common with biological neural networks." Thus, ANNs can be viewed as the parallel interconnection of many simple elements known as neurons (also referred to as nodes) (West et al., 1997). ANNs process information by passing signals between neurons along arcs, which are weighted according to the usefulness of the information being sent. As the network is estimated, weights are adjusted so that the useful arcs are strengthened until the network learns to recognize patterns in the data. The objective is to have the network learn these patterns in such a way that they can be generalized and used to classify new data (Fausett, 1994; West et al., 1997). It is the network structure (or architecture) that gives rise to the functional form of the resulting flexible-regression function.

A neuron takes inputs, x_k, weighted by a parameter, w_k, from K other neurons, aggregates them to obtain a single value, "*net*", and then performs a nonlinear transformation of *net*, $\mathcal{F}(net)$, to produce an individual output, y. Here, $\mathcal{F}(net)$ is termed an "activation function" and is commonly the logistic or hyperbolic tangent function (West et al., 1997). An intercept term can also be added to yield (Fausett, 1994):

$$net = a + \sum_{k=1}^{K} w_k x_k \tag{4.14}$$

and

$$y = \mathcal{F}(net) = \mathcal{F}\left(a + \sum_{k=1}^{K} w_k x_k\right), \tag{4.15}$$

which is depicted in figure 4.1.

At a minimum, ANNs consist of an input layer and an output layer, but hidden layers — layers of neurons between the input and output layers — can be added to approximate highly nonlinear functions. A researcher can think of each hidden layer as a way to reduce the dimensionality of the problem to improve the approximation capabilities of the ANN. Figure 4.2 illustrates the structure of a single-hidden-layer feed-forward ANN. In a single-hidden-layer network, inputs $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,K})$ from the i^{th} observation are introduced to the input layer neurons which send signals $w_{k,h} x_{i,k}$ to each neuron in the hidden layer, where k and h denote the neurons sending and receiving the signal, respectively. Each neuron in the hidden layer aggregates the input signals received to form $net_{i,h}$, which is then transformed using an activation function to obtain an output:

$$y_{i,h} = \mathcal{F}_1\left(net_{i,h}\right), \ \ h = 1, \ldots, H \tag{4.16}$$

where

$$net_{i,h} = a_h + \sum_{k=1}^{K} w_{k,h} x_{i,k} \tag{4.17}$$

and $\mathcal{F}_1(\cdot)$ is the hidden-layer activation function. Each hidden-layer neuron then sends a signal $w_h y_{h,i}$ to the output layer. The output layer sums the signals to obtain $net_i = a + \sum_{h=1}^{H} w_h y_{h,i}$, which is then transformed using a second activation function. The resulting output is given by:

$$y_i = \mathcal{F}_2\left(net_i\right) \tag{4.18}$$

where $\mathcal{F}_2(\cdot)$ is the output-layer transformation function and

$$net_i = \sum_{h=1}^{H} w_h \mathcal{F}_1\left(a_h + \sum_{k=1}^{K} w_{k,h} x_{i,k}\right). \tag{4.19}$$

Assuming a bias term is included at the output layer, the approximation of the conditional mean given by equation 4.9 can be modeled using a single-hidden-layer network and can be represented as (Mehrotra et al., 1997; West et al., 1997):

$$E\left(y_1 \mid \mathbf{X} = \mathbf{x}_i\right) = \mathcal{F}_2\left(a + \sum_{h=1}^{H} w_h \mathcal{F}_1\left(a_h + \sum_{k=1}^{K} w_{k,h} x_{i,k}\right)\right). \tag{4.20}$$

While multiple hidden layers can be considered, only single-hidden-layer networks are examined in this study.

4.3.2 Statistical Theory

Given $E\left(|y_i|^2\right) < \infty$, y_i is square integrable and is a member of $L_2\left(\mathcal{X}, \mu\right)$, the set of all square-integrable real-valued functions (with range $\mathcal{X}$) with respect to a finite, non-negative measure μ, that is absolutely continuous with respect to Lebesque measure (or counting measure in the case of a discrete random variable)[3]. Assuming that $E\left(|x_{i,k}|^2\right) < \infty$ for $k = 1, \ldots, K$, then $\mathbf{x}_i \in L_2(\mathcal{X}, \mu)$ and is treated as a random vector (or set of random variables). In addition, it should be noted that any Borel functions of the elements of $\mathbf{x}_i$ (e.g., $g_j(\mathbf{x}_i)$) are members of $L_2(\mathcal{X}, \mu)$ as well (see Billingsley (2008) or Spanos (1999)). Thus, by the classical projection theorem, $E(y_i \mid \mathbf{X} = \mathbf{x}_i)$ is also a member of $L_2(\mathcal{X}, \mu)$. In this sense, $E(y_i \mid \mathbf{X} = \mathbf{x}_i)$ is the projection of y_i onto the subspace spanned by $\mathbf{x}_i$ (and/or the Borel functions of $\mathbf{x}_i$) (Luenberger, 1969; Small & McLeish, 2011). Leshno, Lin, Pinkus, and Schocken (1993) state if μ is given as above and $\mathcal{X}$ is compact, then the set of all single-hidden-layer feed-forward ANNs with linear output-neuron activation functions are dense in $L_2(\mathcal{X}, \mu)$ with respect to the L_2 metric, as long as the hidden-layer activation functions are not almost-everywhere polynomials, locally bounded, and discontinuous only on a set of measure zero, which includes ANNs specified using sigmoid hidden-layer functions (see Fine (2006), as well). Ripley (1994) noted that this can be extended to a single-hidden-layer FFBANN with a logistic-activation function (or potentially any other appropriate cdf) in the output layer as long as the values taken by the network are bounded away from 0 and 1. These results imply that a FFBANN can be used to approximate $E(y_i \mid \mathbf{X} = \mathbf{x}_i)$ in $L_2(\mathcal{X}, \mu)$ and can be regarded as a flexible functional form.

The approximation results allow FFBANNs to be viewed as SNP alternatives to the binary logit and probit models. If the researcher is concerned about potential misspecification of equation 4.9, then the modeler may wish to approximate $E(y_i \mid \mathbf{X} = \mathbf{x}_i)$. Using a single-hidden-layer FFBANN (with bias), gives rise to the following SNP regression function:

$$y_i = \mathcal{F}_2\left(a + \sum_{h=1}^{H} w_h \mathcal{F}_1\left(a_h + \sum_{k=1}^{K} w_{k,h} x_{i,k}\right)\right) + u_i \tag{4.21}$$

where $y_1 \sim \mathit{Bernoulli}(p)$ with variance $p(1-p)$. Then, for example, a single-hidden-layer

[3]Note that if μ is absolutely continuous with respect to Lebesque measure, then it implies that μ has a density function by the Radon-Nikodym Theorem (Billingsley, 2008).

FFBANN with a single output neuron and logistic-activation function is

$$y_i = \left\{ 1 + \exp \left(- \left[a + \sum_{h=1}^{H} w_h \mathcal{F}_1 \left(a_h + \sum_{k=1}^{K} w_{k,h} x_{i,k} \right) \right] \right) \right\}^{-1} + u_i \tag{4.22}$$

or

$$y_i = \left[1 + \exp \left(-net_i \right) \right]^{-1} + u_i \tag{4.23}$$

where $net_i = a + \sum_h w_h \mathcal{F}_1 \left(a_h + \sum_k w_{k,h} x_{i,k} \right)$ is a single-hidden-layer FFBANN with a single output neuron and a linear-activation function. According to Hornik et al. (1989), such a network can approximate any continuous function uniformly. Thus, it can be interpreted as uniformly approximating the index or predictor function of a logistic-regression model. That is, the single-hidden-layer FFBANN $net_i \left(\mathbf{x}_i; \mathbf{w} \right)$ in equation 4.23 can be viewed as an approximation for the index function given by $\mathcal{I} \left(\mathbf{x}_i; \boldsymbol{\beta} \right)$ in equation 4.9.

4.3.3 Estimation

A particular concern during estimation (or training) is the question of how well the FFBANN performs in classifying input patterns that were not used to estimate the network, or generalizability. This issue arises due to the fear that the network will be over-fit. Fine (2006, p. 155) states that "fitting too closely to the training set means fitting to the noise [in the data] as well and thereby doing less well on new inputs that will have noise independent of that found in the training set." To avoid over-fitting, a validation data set that is independent of the training data set is constructed or set aside from the original sample (Principe et al., 2000). The validation set is then used in conjunction with a stopping rule based on an out-of-sample performance measure to terminate training. This technique is known as cross-validation. Two commonly used measures are to terminate when — after a pre-specified number of iterations — either (1) the validation-data mean square error (MSE) does not decrease or (2) the number of input patterns correctly classified does not increase (Fine, 2006; Kastens & Featherstone, 1996). Using these two stopping rules will result in estimating connection weights less precisely and thereby not iterating to a convergent solution for the training data set. Such stopping points are desirable if the modeler wants to achieve better generalization and avoid over-fitting (Kastens & Featherstone, 1996).

The MSE stopping rule amounts to estimating the ANN using nonlinear least squares (NLS) [4], but from a cross-validation perspective to avoid over-fitting the network. The objective is to minimize the MSE on the training data, i.e.:

$$\min E\left(\cdot\right) = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2, \tag{4.24}$$

where y_i is the dependent-variable observation from the training data set and $\hat{y}_i$ is the fitted or predicted value from the ANN Mehrotra et al. (1997). Finding the parameter values — i.e., weights and bias terms — that minimize $E\left(\cdot\right)$ is an unconstrained optimization problem. If $E\left(\cdot\right)$ is differentiable, parameters can be updated recursively during estimation using the chain rule of differentiation in a process known as back-propagation.

Note that in the case of no hidden layers and a logistic-activation function, the network is simply a numerical estimation of a standard logistic-regression model, and in general, minimizing the MSE to estimate the parameters is a purely NLS problem. White (1989) and Kuan and White (1994) establish the necessary conditions for consistency and asymptotic normality of the NLS estimator for the network parameters.

An additional consideration when using ANNs, as with any numerical optimization, is that changes in starting points can affect the network's performance and parameter estimates. Additionally, certain learning machines, including ANNs, have been found to be unstable (Breiman, 1996). For unstable procedures, small changes in the training data set can lead to large changes in estimation results (Breiman, 1996). To address such issues, Breiman (1996) suggests "bootstrap bagging", or simply bagging. Within the ANN framework, bagging refers to multiple estimations of the selected network architecture with randomized subsets of the

[4] Another commonly used fitting criterion is the Kullback-Leibler criterion. When this criterion is used, parameters are essentially estimated via maximum likelihood estimation (Bergtold, 2004).

data used as the training set for each estimation. The bagging process can also provide a distribution of parameter estimates and functions of those parameters, e.g., marginal effects and WTP. Thus, bagging allows for statistical inferencing, which from a single estimation of a neural network is not straightforward or likely reliable.

4.3.4 Marginal Effects

Most ANNs are used for purely predictive or classification purposes. There has not been much work in the applied literature on the use of ANNs for substantive inference. Of particular interest to economists and other social scientists is the marginal effect of an explanatory variable on the likelihood of an outcome, especially in the context of discrete choice models. As with the logit or probit models, the marginal effect from ANNs associated with a specific explanatory variable is generally not equal to a single parameter value[5]. Consistent with the "black box" nature of ANN parameter values, generally little can be inferred from the parameters about marginal effects. This contrasts with the logit and probit models for example, where the sign of a marginal effect can often be obtained from the sign on a specific parameter. Thus, if one is interested in how changes in explanatory variables impact choice probabilities given by an ANN, an analytical derivation of the marginal effects provides a solution. While not seen in the literature, the derivation of these marginal effects for discrete choice models involving ANNs is relatively straightforward using the chain rule. This sub-section derives marginal effects for single hidden layer FFBANNs for binary choice processes. These derivations provide a unique contribution to the applied literature and further establishes the potential applicability of these models for discrete choice modeling, as well as analyzing CV and stated-choice survey data.

When $\mathcal{F}_1(\cdot)$ — the activation function in the hidden layer — is the logistic cdf, the

[5] In the case of an ANN with no hidden layers and linear activation function in the output neuron, the marginal effect for a given explanatory variable will be a single parameter value.

network output can be written as:

$$\hat{y}_i = \left\{ 1 + \exp\left[- \left(a + \sum_{h=1}^{H} w_h \left\{ 1 + \exp\left[- \left(a_h + \sum_{k=1}^{K} w_{k,h} x_{i,k} \right) \right] \right\}^{-1} \right) \right] \right\}^{-1}. \qquad (4.25)$$

When the explanatory variable of interest, say x_j, is binary, the marginal effect for the i^{th} individual can be calculated as the discrete difference between the two possible states:

$$ME_{i,j} = \hat{y}_{i|1} - \hat{y}_{i|0} \qquad (4.26)$$

where $\hat{y}_{i|1}$ represents the network estimate for individual i when $x_{i,j} = 1$ and $\hat{y}_{i|0}$ represents estimate when $x_{i,j} = 0$. If the explanatory variable of interest is continuous, the marginal effect for the i^{th} individual becomes the partial derivative of equation 4.25 with respect to the variables:

$$ME_{i,j} = \frac{\partial \hat{y}_i}{\partial x_{i,j}} = \frac{\partial \left[\left\{ 1 + \exp\left[- \left(a + \sum_{h=1}^{H} w_h \left\{ 1 + \exp\left[- \left(a_h + \sum_{k=1}^{K} w_{k,h} x_{i,k} \right) \right] \right\}^{-1} \right) \right] \right\}^{-1} \right]}{\partial x_{i,j}}. \qquad (4.27)$$

Applying the chain rule yields:

$$ME_{i,j} = \left[\mathcal{F}_2 \left(net_i \right) \right] \left[1 - \mathcal{F}_2 \left(net_i \right) \right] \sum_{h=1}^{H} w_h w_{j,h} \left[\mathcal{F}_1 \left(net_{i,h} \right) \right] \left[1 - \mathcal{F}_1 \left(net_{i,h} \right) \right] \qquad (4.28)$$

or

$$\begin{aligned} ME_{i,j} = \; & \left[1 + \exp\left(-net_i \right) \right]^{-1} \left\{ 1 - \left[1 + \exp\left(-net_i \right) \right]^{-1} \right\} \times \\ & \textstyle\sum_{h=1}^{H} w_h w_{j,h} \left[1 + \exp\left(-net_{i,h} \right) \right]^{-1} \left(1 - \left[1 + \exp\left(-net_{i,h} \right) \right]^{-1} \right) \end{aligned} \qquad (4.29)$$

where $net_{i,h} = a_h + \sum_{k=1}^{K} w_{k,h} x_{i,k}$ and $net_i = a + \sum_{h=1}^{H} w_h \mathcal{F}_1 \left(a_h + \sum_{k=1}^{K} w_{k,h} x_{i,k} \right)$.

Another commonly used sigmoid activation function is the hyperbolic tangent. Replacing the hidden-layer activation functions, $\mathcal{F}_1$, with the hyperbolic tangent, equation 4.25 becomes

$$\hat{y}_i = \left[1 + \exp\left(-\left\{a + \sum_{h=1}^{H} w_h \left(2\left\{1 + \exp\left[-2\left(a_h + \sum_{k=1}^{K} w_{k,h} x_{i,k}\right)\right]\right\} - 1\right)\right\}\right)\right]^{-1} \tag{4.30}$$

For binary explanatory variables, marginal effects can be calculated using equation 4.26. For continuous variables, marginal effects can again be obtained by applying the chain rule to equation 4.30, which yields:

$$ME_{i,j} = 4\mathcal{F}_2(net_i)\left[1 - \mathcal{F}_2(net_i)\right] \sum_{h=1}^{H} w_h w_{j,h} \left[1 + \exp(-2net_{i,h})\right]^{-2} \exp(-2net_{i,h}) \tag{4.31}$$

or

$$\begin{aligned} ME_{i,j} = \; & 4\left[1 + \exp(-net_i)\right]^{-1}\left\{1 - \left[1 + \exp(-net_i)\right]^{-1}\right\} \times \\ & \textstyle\sum_{h=1}^{H} w_h w_{j,h} \left[1 + \exp(-2net_{i,h})\right]^{-2} \exp(-2net_{i,h}). \end{aligned} \tag{4.32}$$

4.3.5 Willingness-to-Pay

The importance of marginal effects in modeling behavior processes is hard to overstate. However, as the name suggests, marginal effects are of secondary importance in CV studies where the primary interest lies in the valuation of environmental amenities or product attributes. Commonly, this is accomplished through estimates of willingness-to-pay for the amenity or attribute in question. Being able to estimate how individuals value these amenities or attributes can have significant consequences in terms of assessing political or economic feasibility. This paper uses median WTP as opposed to mean WTP and offers three justifications for this choice. First, when the political feasibility of a policy is in question, it is more important to identify a level of cost for which a majority of a population will (or will not) support it. Median WTP satisfies this need. Second, median WTP is likely to be more robust to noise in the data compared to mean WTP (Hanemann, 1984). Finally, median WTP has the advantage that the researcher does not need to find the cost such that the probability of adoption is (essentially) zero (Creel & Loomis, 1997).

As described in section 4.2, maximum WTP is taken to be the amount $\$C$ such that the individual is indifferent between responding "yes" or "no." That is, an individual's maximum WTP is the amount $\$C$ such that $P\left[y_i = 1 \mid \mathbf{X} = \mathbf{x}_i, C\right] = P\left[y_i = 0 \mid \mathbf{X} = \mathbf{x}_i, C\right] = 0.5$. For standard logit and probit models, this amount can be found by setting

$$\Delta v = v_1\left(q_1, I - C, \mathbf{s}\right) - v_0\left(q_0, I, \mathbf{s}\right) = 0 \tag{4.33}$$

and solving for C, since $F_\eta\left(0\right) = 0.5$ when $F_\eta\left(\cdot\right)$ is the logistic or standard normal cdf (Cooper, 2002). For FFBANNs, WTP can be found via a line-search procedure (see section 4.4.2 and figure 4.3) to find the amount $\$C$ such that

$$\mathcal{F}_2\left(a + \sum_{h=1}^{H} w_h \mathcal{F}_1\left(a_h + \sum_{k=1}^{K} w_{k,h} x_{i,k}\right)\right) = 0.5 \tag{4.34}$$

where C is included among $\mathbf{x}_i$. The algorithm used in this paper is based on the golden search line-search procedure presented by Bazaraa, Sherali, and Shetty (2013).

4.4 Empirical Applications

This study examines the performance and estimates marginal effects and WTP for various FFBANN models using two different case studies. The first application uses data collected by a research team for a study examining water issues in the Smoky Hill Watershed in western Kansas. The second makes use of publicly available data from a study conducted in the Layawan Watershed, Philippines, by Calderon et al. (2012). The data sets, empirical procedures, and results are described in the sections that follow. Instances from both applications highlight the potential bias in inferences in binary-discrete-choice models from misspecification and how ANNs as a SNP method may overcome it. A similar result is seen with respect to WTP in the second application.

4.4.1 Water Consumption and Drought Occurrence in Communities of the Smoky Hill Watershed

The Smoky Hill Watershed encompasses approximately 20,000 square miles that stretch from central Kansas to eastern Colorado. Two sub-watersheds from the Smoky Hill combine to form a study area that is comprised of roughly 2,440 square miles in central Kansas. Data for this study were collected via survey to examine community members' knowledge and beliefs about the local environment and their willingness to adopt and vote for environmental and conservation policies. Data statistics and descriptions for this study are presented in table 4.1.

Community members from the study region and surrounding counties were surveyed in two phases. In the first phase — July and August of 2015 — surveys were administered to visitors at county fairs in communities within and around the study region. Respondents were randomly selected for participation and screened based on whether they were at least 18 years of age and resided in counties within and around the watershed. Those who met these criteria were offered a $15 payment for their participation in the survey. A total of 679 surveys were handed out at five county fair venues. Of these, 558 were completed and returned for a response rate of 82.2%. The second phase — September to December of 2015 — was conducted by mailing versions of the survey to two different groups within the watershed: farmers and non-farmer community members. Farmers were selected from a contact list obtained from FarmMarketID and who had responded to land use surveys administered by the team in past years. A non-farmer community member list was obtained from directmail.com. From this list, a random sample was pulled for each county proportional to that county's share of the region's population. Mail surveys to farmers and non-farmers included a $2 incentive. A total of 474 farmer and 2,526 non-farmer surveys were mailed. From the farmer sample, 113 were returned completed (response rate of 26.5%) and from the non-farmer sample, 717 were returned (response rate of 31.4%). Combining the two phases, a total of 1,388 surveys were completed for an effective response rate of 40.8% overall. For the purposes of this study, due to incomplete responses, 1,007 surveys were usable for analysis.

This study focuses on community member responses when faced with the option of paying a percentage increase in their monthly water bill in order to maintain current water usage levels during times of drought. The percentage increase varied across survey versions from 1% to 100%. Using the percentage increase, the monthly payment faced by an individual was calculated as

$$AMOUNT_i = \text{percent}_i \times (\text{monthly payment}_i)\,, \tag{4.35}$$

where the monthly payment was based on a survey question asking for the respondent's average monthly water bill. Descriptions and statistics for the variables in this data set are found in Table 1. This procedure resulted in monthly-water-bill increases that ranged between \$0.05 and \$162.50. A positive WTP in this study would indicate that a respondent is willing to fund policies that could ensure an adequate water supply to maintain current levels of use.

4.4.2 Protection and Management of the Layawan Watershed

Located in the Mt. Malindang Range in the Zamboanga Peninsula, Philippines, the Layawan Watershed is a major rain-catchment area and supplies water to the Misamis Occidental, Zamboanga del Sur, and Zamboanga del Norte provinces (Calderon et al., 2012). The watershed has a total area of roughly 41.3 square miles that is approximately 57.8% forest and 41.3% cropland, with the remaining area devoted to rice paddies or urban development (Calderon et al., 2012). Calderon et al. (2012) surveyed 400 households within the Layawan Watershed to examine WTP to manage and protect the watershed in order to have a sustainable water supply and lessen the impacts of natural disasters. The survey used by Calderon et al. (2012) asked respondents if they would be willing to pay a certain bid amount, ranging from 10 to 200 Philippine pesos (10₱to 200₱) over and above the current water bill, to fund conservation efforts. Data from this study was made publicly available by its authors and can be found on-line. Descriptions and statistics for the variables from Calderon et al. (2012) used in this study can be found in table 4.1.

4.4.3 Models and Methods

Kastens and Featherstone (1996) note the difficulty of choosing a particular FFBANN because the specification of the functional form (i.e., network architecture) and estimation method (i.e., training style) tend to be problem dependent. Decisions concerning the design of the FFBANN include: (*i*) choice of training algorithm, (*ii*) number of hidden layers, (*iii*) number of neurons in each hidden layer, (*iv*) types of activation functions, (*v*) choice of fitting criterion, and (*vi*) choice of stopping rule. Any one choice for (*i*) — (*vi*) is likely to be dependent upon the choices made for the remaining decisions. Thus, the optimal approach to construct a model would be to perform a grid search over all possible combinations of (*i*) — (*vi*), but such an approach is not practical. To make this more manageable, the only model variations examined in this study are those concerning decision (*iii*) – the number of neurons in each hidden layer. In all, ten network architectures were examined for the Smoky Hill Watershed (SH) data and seven architectures were examined for the Layawan Watershed (LW) data. All specifications consisted of one hidden layer where the number of neurons varied from one to ten for SH networks and one to seven for LW networks. The ranges for the number of hidden-layer neurons is based on a rule of thumb proposed by Bergtold (2004) that suggests that the number of neurons in the first hidden layer be less than or equal to the number of inputs. This rule is centered on the notion that the addition of a hidden layer is aimed at decreasing the dimensionality of the problem, thereby increasing the approximation capabilities of the FFBANN (Bergtold, 2004). Each network was estimated using the Broyden-Fletcher-Goldfarb-Shanno algorithm (decision [*i*]), one hidden layer (decision [*ii*]) with logistic-activation functions in the hidden and output layers (decision [*iv*]), and were fit based on MSE (decision [*v*]). The stopping rule is the same as that described in section 4.3.3. Decisions (i), (ii), (iv) and (vi) were based on research by Bergtold (2004).

To address the issue of instability noted by Breiman (1996), the bagging procedure was employed. The procedure randomly generated 1,000 partitions of the original datasets with each partition being further divided: 80% for training (estimation) and 20% for validation. Additionally, to prevent starting point bias in the parameter estimates and obtain the best local optima during estimation, for each partition, the network was estimated 100 times using random initial values for the parameters. The initialization was performed using the procedure developed by Nguyen and Widrow (1990) that chooses weights so that the hidden neurons will be able to learn the input-output mapping more readily, decreasing estimation time (Fausett, 1994). Results from the best performing network — based on MSE on the validation dataset — from the set of 100 initializations were kept for further use (e.g., estimation of marginal effects and WTP). This procedure was done for each of the SH and LW network specifications.

For the empirical comparisons, binary logit and probit models, as well as the SNP approach developed by Klein and Spady (1993) (KS) were estimated. The KS approach maximizes a pseudo-log-likelihood function that uses nonparametric kernel estimators to approximate the unknown probability function. This procedure was used as it is available in LIMDEP. Logit, probit, and KS models were estimated using LIMDEP (W. Greene, 2012), while ANNs were estimated in MATLAB[6]. The ANNs, logit, probit, and KS models were estimated using the same set of explanatory variables (table 4.1) for the SH data. For the LW data, the logit and probit models included interaction terms between the proposed water bill increase ($AMOUNT$) and variables indicating (1) whether the payment scheme was mandatory or voluntary ($PAYSCH$) and (2) whether all water users or only domestic water users would be subject to the increase ($IPAYEE$). Because respondents belong to a specific group — domestic or non-domestic water users — their willingness to support the proposal is likely influenced by whether or not they bear the burden of the cost. Similarly, whether or not a payment is mandatory or voluntary should impact the cost level at which

[6]Many different econometric software packages provide procedures or add-ins for estimating ANNs (e.g., MATLAB, R, SAS, STATA, EXCEL).

individuals would be willing to support the proposal. *Ceteris paribus*, it is expected that if payment is voluntary, individuals would be willing to support the proposal at higher costs due to free riding. Including these interaction terms allows the marginal effect associated with $AMOUNT$ to be different based on (1) whether payment is voluntary and (2) whether all or only some respondents bear the cost.

Marginal effects for FFBANN models were calculated using the methods outlined in section 4.3. For each network specification, marginal effects were computed at the individual level for the best-performing networks estimated from the 1,000 data partitions. Marginal effects were then averaged across individuals and data partitions to yield a "bagged average" marginal effect. For the j^{th} explanatory variable, the bagged marginal effect can be represented as

$$BME_j = \frac{1}{1000}\sum_{r=1}^{1000}\left(\frac{1}{N}\sum_{i=1}^{N} ME_{r,i,j}\right), \tag{4.36}$$

where r denotes the network estimated using the r^{th} data partition and $N = 1,007$ for the SH data set and $N = 399$ for the LW data set. For logit and probit models, the built-in post-estimation command "PARTIAL EFFECTS" was used in LIMDEP, which computes the average marginal effect across observations. This command was also used for the KS models, but for this estimator LIMDEP produces marginal effects calculated at the means of the independent variables. Asymptotic standard errors for FFBANN marginal effects were calculated as the standard deviation of the bagged-average marginal effects. Asymptotic standard errors for the logit, probit, and KS models were estimated using the delta method (W. Greene, 2012).

Following Cooper (2002), for the case where Δv takes the form $\Delta v = \mathbf{x}'\boldsymbol{\beta}$, then letting $\mathbf{x}_i'\boldsymbol{\beta} = \mathbf{z}_i'\boldsymbol{\tau} + \phi AMOUNT_i$, where $AMOUNT_i$ represents the payment amount as in equation 4.35 for the SH data, the median WTP for the i^{th} individual for the SH logit and probit models can be calculated as

$$WTP_i = \frac{-\mathbf{z}_i'\boldsymbol{\tau}}{\phi}. \tag{4.37}$$

In the case of the LW logit and probit models,

$$\mathbf{x}_i'\boldsymbol{\beta} = \mathbf{z}_i'\boldsymbol{\tau} + \phi_1 AMOUNT_i + \phi_2 AMOUNT_i \times PAYSCH_i + \phi_3 AMOUNT_i \times IPAYEE_i \tag{4.38}$$

due to the inclusion of interaction terms. Thus, for the LW logit and probit models, equation 4.37 becomes

$$WTP_i = \frac{\mathbf{z}_i'\boldsymbol{\tau}}{\phi_1 + \phi_2 PAYSCH_i + \phi_3 IPAYEE_i}. \tag{4.39}$$

To find $\$C$ from equation 4.33 for the FFBANNs and the KS model, a modified golden-search procedure from Bergtold (2004) was used (presented in figure 4.1). The procedure searches in a closed interval on the real line where the end points of the interval represent the upper and lower bounds of a respondent's WTP. The intervals were set at $[\$0, \$200]$ for the SH models and $[0\text{P}, 225\text{P}]$ for the LW models. This process is done for each individual in the entire dataset for all 1,000 retained networks to obtain bagged estimates and standard errors for each network specification. The delta method was used to obtain standard errors for the logit and probit models and the method of Krinsky and Robb (1986) was used to obtain standard errors for the KS WTP estimate. For the Krinsky and Robb (1986) procedure, 5,000 WTP estimates are calculated using parameter values drawn from a multivariate-normal distribution based on the original parameter values and covariance matrix from the KS model.

4.5 Results

As motivation for the use of FFBANNs as an alternative to the logit and probit models, misspecification tests were conducted for these models in both case studies. To test the null hypothesis of a linear index function a Ramsey-type RESET test was used based on Bergtold et al. (2010). The RESET test can be used to indicate if higher order terms should be included in the logit and probit index functions. Results from these tests, shown in table 4.2 rejected the null hypotheses of the traditional linear index functions in the logit and probit models for both case studies. In general, results from the different approaches are more similar in the Smoky Hill Watershed application. There were, however, a couple of important differences seen in the marginal effects. The Layawan Watershed application better demonstrates the potential impacts of misspecification, as differences are seen in marginal effects and willingness-to-pay estimates.

4.5.1 Model Fit Comparisons

Model fit statistics for all the models estimated are presented in table 4.3 and figures 4.4 and 4.5. Comparisons between models were based on the percent of outcomes correctly classified (PCC) and mean square error (MSE). The MSE measures represent errors across the entire dataset, which in the case of the FFBANNs includes both the training and validation subsets. Estimated marginal effects and WTP measures for each of the case studies are presented in tables 4.4 to 4.6. Model names in the tables are preceded by either "SH" or "LW" to indicate for which case study the model was estimated. Names for estimated ANN models include a number on the end to indicate the number of neurons in the hidden layer. For example, SH_ANN7 indicates an artificial neural network with seven hidden-layer neurons that was estimated using the Smoky Hill Watershed data.

SH Case Study

On average, the estimated FFBANNs performed better than the logit, probit, and KS models in terms of both PCC and MSE. Because the bagging procedure was used for the ANNs, minimum, maximum, and average values for the fit statistics are provided in the table. Only one estimation was done for the logit, probit, and KS models, so the average, minimum, and maximum values are the same. For the SH case study, the logit, probit, and KS models correctly classified 74.4% to 74.7% of the observations. The lowest average PCC on the SH data across ANNs was 76.4% for SH_ANN10 and the highest was 77.5% for SH_ANN4, and on average PCC was higher for the ANNs by 2 to 3%. The SH_KS PCC, while similar to the logit and probit performance, may be misleading. The KS estimator predicted a vote of "no" for 1,006 of the 1,007 observations. Thus, the KS PCC is essentially just the percentage of individuals who responded "no" on the survey.

With respect to MSE for the SH case study, the average MSEs produced by the ANNs were lower than those of the logit, probit, and KS in all cases by 5 to 13%. SH_ANN4 and SH_ANN5 produced the lowest average MSE scores at 0.1646, while the highest was SH_KS at 0.1887. SH_Logit and SH_Probit produced similar results of 0.1773 and 0.1783, respectively.

LW Case Study

With the LW case study, the logit and probit models again produced similar results with PCCs of 73.9% and 72.9%, respectively. MSE was 0.1741 for LW_Logit and 0.1758 for LW_Probit. The KS model performed considerably worse with a PCC of 51.1% and MSE of 0.2368. When looking at average PCC and average MSE for the ANNs, all nine ANN specifications outperformed the LW_Logit, LW_Probit, and LW_KS models. The best average PCC was 76.7% in LW_ANN6 while the worst average PCC was 75.2% in LW_ANN3. The lowest average MSE was for LW_ANN6 at 0.1646 and the highest was 0.1730 for LW_ANN3.

The results from both case studies underscore the notion that choosing a network architecture is problem dependent and is often best resolved through trialling multiple designs. However, these results also suggest that even choosing a "wrong" architecture may, on average, produce MSEs and PCCs that are superior to the logit, probit, or KS approaches, providing some reassurance of the robustness of this modeling approach.

4.5.2 Marginal Effects Comparisons

Estimated marginal effects for each case study are presented below and in tables 4.4 and 4.5. To the author's knowledge, marginal effects associated with FFBANN models have not presented elsewhere in the literature and thus are a novel result. The ability to estimate marginal effects — and associated standard errors — for ANNs moves this technique beyond the purely predictive realm towards the ability for statistical inference.

SH Case Study

Estimated marginal effects for the SH case study showed some consistencies across models, but also some important differences. Estimated marginal effects were statistically insignificant across all models for the following variables: respondent's age (*AGE*), whether they hold a bachelor's degree (*COLLEGE*), number of individuals in the household (*HHSIZE*), if they identify racially as white (*WHITE*), if they are aware of the depleting level of the Ogallala Aquifer (*KSCARCITY*) or recent droughts in Kansas (*KDROUGHT*). Whether a respondent had voted in a local election in the last four years (*LOCAL*) was statistically insignificant across all models except for the KS approach, where it was statistically significant at the 1% level. For the KS, the *LOCAL* marginal effect suggests that having voted in a local election increases the probability of accepting the higher water bill. The marginal effect associated with the *GENDER* variable was found to be negative and significant in the logit, probit, and all ANN specifications, indicating that men are less likely than women to pay to maintain their current level of water consumption during drought conditions. The *GENDER* marginal effect was insignificant in the KS model.

Two key areas of contrasts between the ANN models and the logit, probit, and KS models were with respect to the *INCOME* and *AMOUNT* variables. Estimated marginal effects for *INCOME* were positive and significant in the logit and probit models, negative and significant in the KS model, and statistically insignificant in all ANN specifications. While the ANN results may be counter to traditional thinking, it is plausible nonetheless considering that, on average, the annualized cost of the water bill increase was about 1% of a household's income. Given the relatively minor share of household income represented by the cost increases, it may be that decisions on this question are being driven more by cultural factors or personal beliefs, e.g., a desire to conserve water in times of drought. The estimated marginal effect for $AMOUNT$ was negative and significant in the logit, probit, and KS models. This was the expected result, indicating that as the cost of maintaining your current level of water consumption increases, individuals are less likely to accept that cost in order to maintain water usage levels. In contrast, this marginal effect was not found to be statistically significant in seven of the ten ANN specifications. For those ANN specifications where the marginal effect was statistically significant, it was found to be negative. It is also possible that the contrasting findings for the $INCOME$ and $AMOUNT$ marginal effects are the result of a misspecified index function. Whatever the underlying cause for the differences, policy decisions based on results from the logit, probit, or KS models could lead to the enactment of policies that may have no significant impact or fail to pass a public referendum.

LW Case Study

With the LW case study, a common conclusion was reached for only two of the seven estimated marginal effects. The first point of agreement was with respect to the proposed increase in the monthly water bill ($AMOUNT$), for which the estimated marginal effect was negative and significant across all models. Second was the marginal effect associated with whether the payment was mandatory or voluntary ($PAYSCH$) that was found to be negative but statistically insignificant in all models. Whether the respondent was male or female ($GENDER$) was statistically insignificant in all models. A respondent's age (AGE) did not generally have a statistically significant marginal effect, except for the LW_ANN1 and LW_ANN2 models. In these models, the estimated marginal effect was positive and significant, indicating that older individuals are more likely to accept the proposal. Respondents' monthly household income ($INCOME$) had a positive and significant marginal effect in all models except for LW_ANN1.

Estimated marginal effects for the two remaining variables were less consistent. Whether all water users or only domestic users were subject to the water bill increase (*IPAYEE*) was statistically insignificant marginal effect in the logit, probit, KS, and LW_ANN7 models. In the remaining ANN specifications, this marginal effect was statistically significant and negative, suggesting that if all water users were subject to the increase, respondents were less likely to vote in favor of the proposal. The final marginal effect is associated with the variable *COLLEGE* that is equal to 1 if the respondent had a bachelor's, master's, or vocational degree. Positive and significant marginal effects were found for this variable in six of the seven ANN specifications, with LW_ANN3 as the only exception. For these models, this implies that individuals who hold one of these degrees are more likely to accept the proposal. This marginal effect was not statistically significant in the LW_Logit, LW_Probit, or LW_KS models.

Particularly interesting with the LW marginal effects is the consistency — for the most part — between the neural networks and the logit and probit models with respect to *AMOUNT* and *PAYSCH*. As noted, all models were in agreement with respect to sign and statistical significance. For *AMOUNT* (all statistically significant) there were some differences, however, in terms of the magnitudes. The largest (in absolute value) estimated marginal effect was for LW_Probit while the smallest was for LW_KS. Magnitudes for the ANN marginal effects were consistent across specifications. On average, *AMOUNT* marginal effects from the ANNs were about 37% of those from LW_KS and 63% of those from LW_Probit. For the logit and probit models, the estimated marginal effects for these variables were the result of an index function that had to be specified to include the interaction terms between *AMOUNT* and *PAYSCH*. With the neural networks, however, there is no need for the researcher to create and explicitly include these interaction terms as the ANN is a flexible functional form and implicitly takes these interactions into account.

4.5.3 Median Willingness-to-Pay

Median willingness-to-pay statistics are presented in table 4.6. Average median WTP estimates were calculated using the methods described in section 4.4.

SH Case Study

For the SH dataset, there was some discrepancy in the size of WTP estimates, ranging from \$0.06 in the KS model up to \$3.39 in SH_ANN6. In general, WTP was larger in the neural networks (bagged averages from \$1.35 to \$3.39) than the logit (\$0.10), probit (\$0.09), or KS (\$0.06) models. However, WTP was not statistically significant (up to the 10% level) for any of the models estimated. These results suggest that individuals in the Smoky Hill Watershed would not be willing to pay an increase in their monthly water bill in order to maintain water usage levels during times of drought. One possible reason for this result is that individuals regard maintaining water-use levels during these times as irresponsible, and thus would opt to cut back, perhaps even if no cost was involved. Second, it may be that most individuals feel they could cut back even if they would prefer to keep levels the same. In this case, individuals would be more willing to, for example, reduce lawn watering or shower times rather than face an increase in payments. The finding of no statistical significance across all models is highly significant for policy considerations. Based on these results, policy makers in the Smoky Hill Watershed may be better served during times of drought to impose water restrictions rather than allowing for maintained usage at higher costs. This type of policy approach would not only conserve more water (it is likely some users would maintain current levels under increased costs), but also put policy makers in a position of harmony with the values of community.

LW Case Study

In contrast to the SH models, estimated WTP was statistically significant across all models for the LW data. The smallest WTP came from the KS model at ₱25.15 while the largest was from the probit model at ₱66.02. WTP in the logit model was slightly smaller than the probit at ₱63.24. Falling in between the KS and logit models were the ANNs, which ranged from ₱47.05 in LW_ANN1 to ₱56.60 in LW_ANN4. While the differences may seem small on first glance (₱1.00 ≈ \$0.02), back-of-the-envelope calculations indicate they could still have important policy ramifications. The largest difference between a neural network and either the logit or probit was ₱18.97 (LW_ANN1 and LW_Probit) whereas the smallest was ₱6.64 (LW_ANN4 and LW_Logit). Turning these monthly costs into annual costs yields a difference range of between ₱79.64 and ₱227.58 (\$1.72 to \$4.92). Extending this range to the 4,773 households within the watershed boundary (Calderon et al., 2012) then yields a range of between roughly ₱380 thousand and ₱1.09 million, or between \$23.5 thousand and \$8.2 thousand annually. These differences could play an important role when deciding the political and financial feasibility of such policies. The higher WTP from the logit and probit models may be due to an upward bias resulting from the improperly specified index functions that were indicated by the RESET tests.

4.6 Conclusions

Traditional methods such as logit and probit estimation for CV models are subject to potential misspecification of the index function. While the linear index function of the logit and probit models may be statistically adequate in some situations, Kay and Little (1987) show that the conditions necessary for this adequacy are somewhat stringent. If a misspecified logit or probit model is used as the basis for policy decisions, the resulting policies may have little to no impact if enacted or fail to even be enacted if put to a public vote. Feed-forward back-propagation artificial neural networks (FFBANN) provide a semi-nonparametric alternative to these traditional approaches that can be used to avoid potential misspecification and subsequent ramifications.

This paper used case studies from the Smoky Hill Watershed (SH) in Kansas and the

Layawan Watershed (LW) in the Philippines to demonstrate the potential for FFBANNs as alternatives to traditional logit and probit estimation. Both case studies examined respondents' willingness to support the provision of environmental services, e.g., water supplies, through increases in monthly water bills. Between the SH and LW datasets, a total of 17 different networks were trained. In each case study, ANNs were compared to each other and also to the logit, probit, and the semi-nonparametric estimator of Klein and Spady (1993). Misspecification tests for the logit and probit models in both case studies indicated that the traditional linear (in parameters and explanatory variables) index functions were not appropriate. Comparisons were made with respect to the percent of the dependent variables correctly classified (PCC), mean squared error (MSE), marginal effects estimates, and median willingness-to-pay (WTP) measures. The derivation of the ANN marginal effects and the estimation of both marginal effects, WTP, and associated standard errors for the neural networks provides a novel contribution to the literature and helps to remove some of the "black box" stigma from ANNs by allowing for meaningful insights and statistical inference.

Comparing model fit between the different approaches suggests that FFBANNs used in conjunction with bagging are a viable if not preferable alternative to the logit, probit, and KS estimators. The bagged-average PCC and MSE for the ANNs indicate a greater ability to correctly classify and provide a better model fit. On average, PCC was higher in the individual ANNs than in the logit or probit models by anywhere from roughly 1% to 4%. With the SH data, the MSE in the ANNs averaged about 93% of the logit and probit MSEs and about 78% of that from the KS model. Similarly for the LW data, MSE for the ANNs averaged 96% of the logit and probit and only 71% that of the KS approach. The ANNs outperformed the KS estimator in both datasets with respect to both MSE and PCC, including the misleading KS PCC with the SH data.

Estimated marginal effects saw varying degrees of agreement across the datasets and variables. One notable difference was the marginal effect associated with the proposed cost to respondents using the SH data. This estimate was negative and statistically significant in the logit, probit and KS models but was not statistically significant in seven of ten ANNs. The lack of statistical significance for all WTP estimates using SH data may lend more credibility to the marginal effect on cost from the neural networks. A second interesting result was seen in the LW models. Marginal effects results from this data suggest that the neural networks internally capture interactions between variables without the researcher explicitly creating these interactions. Thus, if a researcher is concerned about potential misspecification, they may want to consider the use of FFBANNs in conjunction with bagging.

Median WTP for the LW data was statistically significant in all models. Average LW WTP estimates from the ANNs were only about 83% of the logit estimate (₱63.24) and 80% of the probit estimate (₱66.02). The KS WTP estimate of ₱25.15 was about one half the ANN estimates on average. If the difference between the logit/probit and the ANNs is due the misspecification in the logit and probit models, this would imply that these traditional approaches are biasing median WTP upward by 20% to 25% in these studies. Based on the results from the performance measures by the KS approach, there is little reason to believe its WTP estimate is accurate. Rather, it is more likely the KS median WTP is biased down in this case. Whether biased up or down, misspecification may harm policymakers' abilities to make informed decisions.

Based on the results of this study, if a researcher is concerned about misspecification in the logit or probit models, they should consider using feed-forward back-propagation artificial neural networks along with the bagging procedure. In fact, other than an increase in computer run time, it is hard to make a case for the other approaches examined in this study over the FFBANNs. The neural networks proved to be superior in terms of fit and predictive capabilities and also appear to avoid issues that could arise from misspecification in the logit, probit, and KS models.

Figures

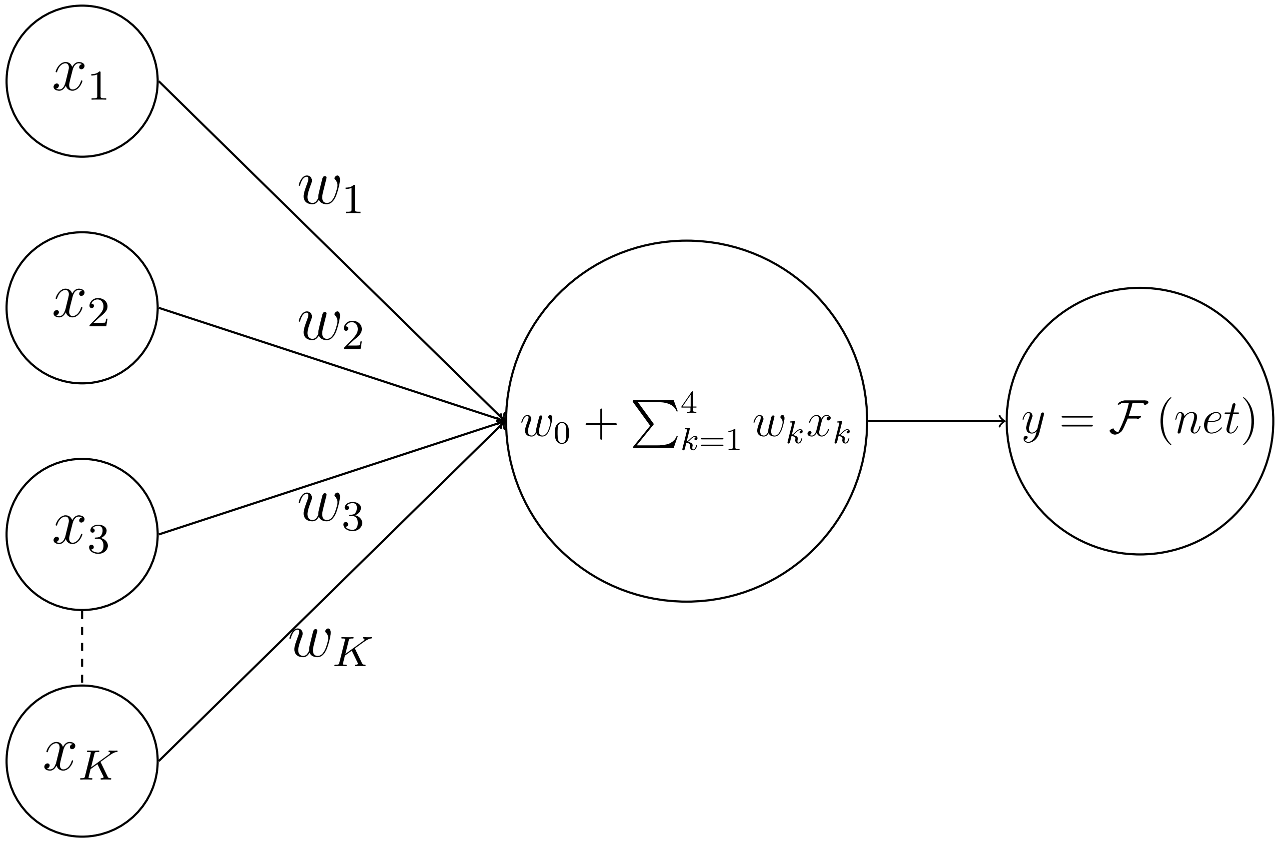

Figure 4.1: Topology of a neuron

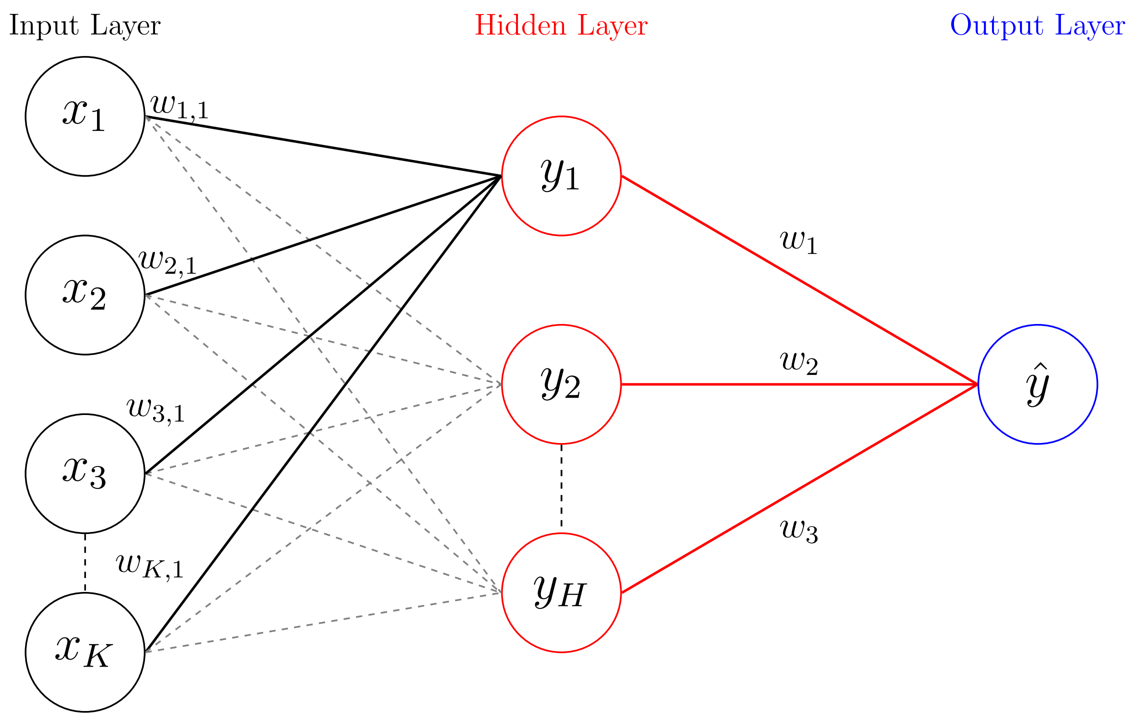

Figure 4.2: Single-Hidden-Layer ANN

Initialization: Determine the initial WTP interval $[a, b]$ on the real line. Choose a tolerance level $\gamma > 0$, which represents how close the algorithm needs to converge to the median WTP to terminate. Set $\lambda = a + (1 - \alpha)(b - a)$ and $\mu = a + \alpha (b - a)$, where $\alpha = 0.618$. For the i^{th} individual, fix the remaining explanatory variables or input values to their current level and calculate $Y_i(a)$, $Y_i(b)$, $Y_i(\lambda)$, $Y_i(\mu)$, where $Y_i(c)$ is the FFBANN output for $C = c$.
Main Step: For the i^{th} individual:

1. If $Y_i(a) \leq 0.5$ or $Y_i(b) > 0.5$, STOP. The median WTP is $C = a$ or $C = b$, respectively. Otherwise, go to step 2.

2. If $|b - a| < \gamma$, STOP. The optimal solution lies in the interval $[a, b]$. Let $C = 0.5(b - a)$. Otherwise, if $Y_i(b) < 0.5$ and $Y_i(\lambda) > 0.5$ go to step 3; if $Y_i(\lambda) < 0.5$ and $Y_i(\mu) > 0.5$ go to step 4; else if $Y_i(\mu) < 0.5$ and $Y_i(a) > 0.5$ go to step 5.

3. Let $b = b$, $a = \lambda$ and $Y_i(a) = Y_i(\lambda)$. Recalculate λ, μ, $Y_i(\lambda)$, and $Y_i(\mu)$ using the new interval $[a, b]$ with the formulas presented in the *Initialization*. Return to step 1.

4. Let $a = \mu$, $b = \lambda$, $Y_i(a) = Y_i(\mu)$, and $Y_i(b) = Y_i(\lambda)$. Recalculate λ, μ, $Y_i(\lambda)$, and $Y_i(\mu)$ using the new interval $[a, b]$ with the formulas presented in the *Initialization*. Return to step 1.

5. Let $a = a$, $b = \mu$, and $Y_i(b) = Y_i(\mu)$. Recalculate λ, μ, $Y_i(\lambda)$, and $Y_i(\mu)$ using the new interval $[a, b]$ with the formulas presented in the *Initialization*. Return to step 1.

Do this for all respondents surveyed to obtain a vector of median WTP values. Once this vector has been obtained, calculate the mean to obtain the median WTP for the entire group of respondents.

Figure 4.3: Median WTP Line-Search Procedure

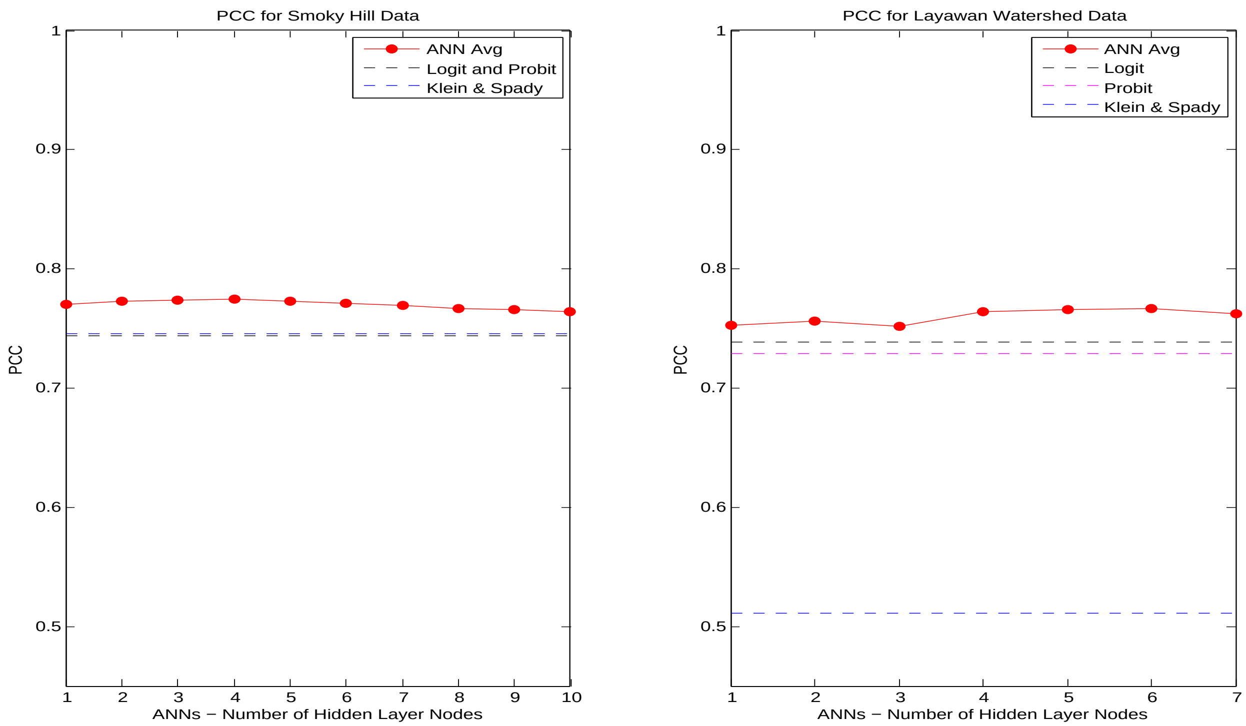

Figure 4.4: Percent correctly classified (PCC) results

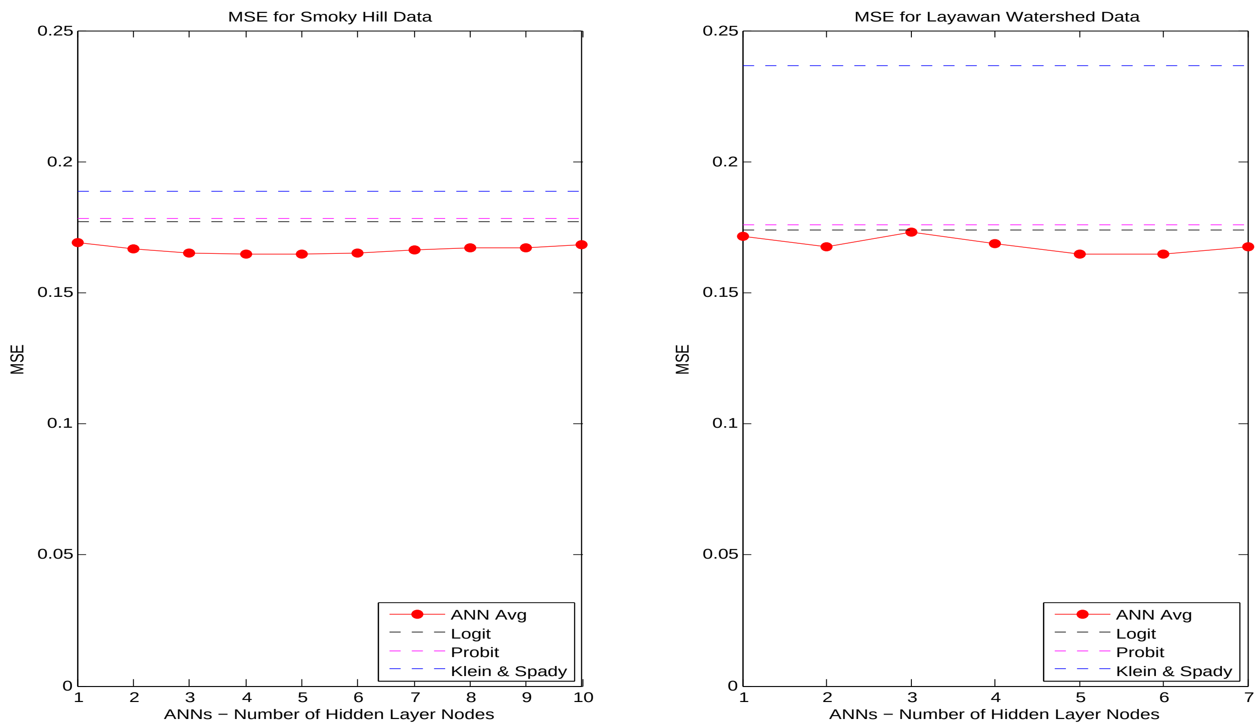

Figure 4.5: Mean square error (MSE) results

Tables

Table 4.1: *Summary data for dependent and explanatory variables*

Variable	Description	Mean	Standard Deviation
Smoky Hill Watershed Data			
VOTE	Dependent Variable. Equal to 1 if respondent would pay the proposed amount to maintain water usage levels during drought.	0.25	0.44
AGE	Respondent's age in years.	50.5	17.1
AMOUNT	Proposed increase in monthly water bill.	\$29.49	\$30.62
COLLEGE	Equal to 1 if respondent has Bachelor's degree or higher, 0 otherwise.	0.32	0.47
HHSIZE	Number of individuals living in household.	2.73	1.53
INCOME	Respondent's income level. Calculated as median of reported income range.	\$65,715	\$63,023
KDROUGHT	Equal to 1 if respondent is aware of recent drought conditions in Kansas.	0.92	0.27
KSCARCE	Equal to 1 if respondent is aware of Ogallala Aquifer depletion.	0.92	0.28
LOCAL	Equal to 1 if responded has voted in a local election in the last four years.	0.72	0.45
GENDER	Equal to 1 for male, 0 for female.	0.51	0.50
WHITE	Equal to 1 if respondent is white, 0 otherwise.	0.78	0.41
Layawan Watershed Data			
VOTE	Dependent variable. Equal to 1 if respondent would pay the proposed amount for the conservation plan.	0.51	0.50
AGE	Respondent's age in years.	48.3	15.4
AMOUNT	Proposed increase in monthly water bill.	₱68.10	₱65.50
COLLEGE	Equal to 1 if respondent's reported education level is College, Vocational, or Master's.	0.25	0.43
INCOME	Total household income per month.	₱8,186	₱10,394
IPAYEE	Equal to 1 if all water users pay, 0 if only domestic water users pay.	0.50	0.50
PAYSCH	Equal to 1 if payment scheme is mandatory, 0 if voluntary.	0.50	0.50
GENDER	Equal to 1 if male, 0 if female.	0.30	0.46

Table 4.2: *Specification test results*

Model	Variable	Coefficient p-value
Smoky Hill		
Logit	$(\mathbf{x}'\boldsymbol{\beta})^2$	0.016
Logit	$(\mathbf{x}'\boldsymbol{\beta})^3$	0.114
Probit	$(\mathbf{x}'\boldsymbol{\beta})^2$	0.041
Probit	$(\mathbf{x}'\boldsymbol{\beta})^3$	0.244
Layawan Watershed		
Logit	$(\mathbf{x}'\boldsymbol{\beta})^2$	0.035
Logit	$(\mathbf{x}'\boldsymbol{\beta})^3$	0.438
Probit	$(\mathbf{x}'\boldsymbol{\beta})^2$	0.025
Probit	$(\mathbf{x}'\boldsymbol{\beta})^3$	0.577

Table 4.3: *Fit statistics across models and datasets*

Model	Min PCC	Max PCC	Avg PCC	Min MSE	Max MSE	Avg MSE
Smoky Hill Watershed						
SH_Logit	74.4%	74.4%	74.4%	0.1773	0.1773	0.1773
SH_Probit	74.4%	74.4%	74.4%	0.1783	0.1783	0.1783
SH_KS	74.7%	74.7%	74.7%	0.1887	0.1887	0.1887
SH_ANN1	73.2%	79.5%	77.0%	0.1607	0.2013	0.1690
SH_ANN2	72.8%	79.9%	77.3%	0.1564	0.2008	0.1667
SH_ANN3	69.4%	80.5%	77.4%	0.1547	0.2078	0.1652
SH_ANN4	72.0%	80.7%	77.5%	0.1488	0.1995	0.1646
SH_ANN5	72.0%	81.3%	77.3%	0.1437	0.2152	0.1646
SH_ANN6	70.8%	80.7%	77.1%	0.1458	0.2097	0.1652
SH_ANN7	71.3%	82.0%	76.9%	0.1393	0.2120	0.1662
SH_ANN8	70.9%	81.3%	76.7%	0.1437	0.2148	0.1672
SH_ANN9	72.0%	81.8%	76.6%	0.1424	0.2164	0.1673
SH_ANN10	71.4%	81.8%	76.4%	0.1424	0.2176	0.1684
Smoky Hill Watershed						
LW_Logit	73.9%	73.9%	73.9%	0.1741	0.1741	0.1741
LW_Probit	72.9%	72.9%	72.9%	0.1758	0.1758	0.1758
LW_KS	51.1%	51.1%	51.1%	0.2368	0.2368	0.2368
LW_ANN1	48.6%	77.9%	75.3%	0.1645	0.2311	0.1715
LW_ANN2	66.4%	83.2%	75.6%	0.1366	0.2405	0.1677
LW_ANN3	62.7%	80.7%	75.2%	0.1504	0.2275	0.1730
LW_ANN4	65.2%	81.7%	76.4%	0.1427	0.2256	0.1687
LW_ANN5	66.2%	81.7%	76.6%	0.1392	0.2354	0.1647
LW_ANN6	63.2%	80.7%	76.7%	0.1438	0.2248	0.1646
LW_ANN7	62.9%	83.7%	76.2%	0.1351	0.2287	0.1675

Table 4.4: *Marginal effects for Smoky Hill Watershed Models*

	Variables									
Model	*AGE*	*AMOUNT*	*COLLEGE*	*GENDER*	*HHSIZE*	*INC*	*KDROUGHT*	*KSCARCITY*	*LOCAL*	*WHITE*
SH_Logit	$1.50E-04$ (0.158)	-0.004^{***} (-5.94)	-0.004 (-0.138)	-0.069^{**} (-2.52)	-0.004 (-0.371)	$3.80E-04^{*}$ (1.73)	0.037 (0.749)	0.059 (1.28)	-0.011 (-0.340)	-0.012 (-0.349)
SH_Probit	$1.50E-04$ (0.160)	-0.003^{***} (-6.39)	-0.007 (-0.242)	-0.068^{**} (-2.49)	-0.003 (-0.318)	$3.80E-04^{*}$ (1.81)	0.038 (0.763)	0.064 (1.37)	-0.012 (-0.372)	-0.012 (-0.365)
SH_KS	$4.40E-04$ (-0.830)	-0.001^{**} (-2.50)	0.018 (0.960)	-0.017 (-0.941)	0.002 (0.401)	$-1.40E-04^{***}$ (-2.14)	$-4.20E-04$ (-0.016)	0.001 (0.042)	0.179^{***} (9.94)	0.025 (—)
SH_ANN1	$4.75E-05$ (-0.008)	-0.026 (-0.086)	0.010 (0.459)	-0.041^{**} (-2.24)	-0.005 (-0.054)	$2.52E-07$ (0.046)	0.001 (0.062)	-0.015 (-0.559)	0.007 (0.306)	-0.006 (-0.273)
SH_ANN2	$6.00E-05$ (0.058)	-0.019^{*} (-1.70)	0.003 (0.097)	-0.048^{**} (-2.56)	-0.004 (-0.490)	$1.61E-08$ (0.050)	0.011 (0.374)	0.015 (0.324)	0.009 (0.342)	-0.008 (-0.321)
SH_ANN3	$9.39E-05$ (0.089)	-0.018 (-1.59)	0.002 (0.057)	-0.053^{***} (-2.59)	-0.003 (-0.470)	$6.41E-08$ (0.201)	0.016 (0.453)	0.026 (0.577)	0.008 (0.300)	-0.008 (-0.364)
SH_ANN4	$1.80E-04$ (0.196)	-0.016 (-1.43)	0.001 (0.021)	-0.055^{**} (-2.49)	-0.004 (-0.494)	$9.39E-08$ (0.283)	0.021 (0.591)	0.041 (0.922)	0.009 (0.320)	-0.009 (-0.421)
SH_ANN5	$2.80E-04$ (0.268)	-0.013 (-1.33)	0.003 (0.11)	-0.059^{**} (-2.48)	-0.004 (-0.503)	$1.68E-07$ (0.533)	0.027 (0.717)	0.047 (0.980)	0.008 (0.253)	-0.010 (-0.378)
SH_ANN6	$3.30E-04$ (0.142)	-0.010 (-1.33)	0.006 (0.24)	-0.063^{***} (-2.89)	-0.004 (-0.232)	$2.52E-07$ (0.429)	0.033 (0.809)	0.050 (1.13)	0.004 (0.147)	-0.011 (-0.424)
SH_ANN7	$4.00E-04$ (0.443)	-0.008 (-1.60)	0.008 (0.31)	-0.064^{***} (-2.64)	-0.005 (-0.603)	$2.48E-07$ (0.943)	0.039 (0.966)	0.045 (1.04)	0.002 (0.076)	-0.012 (-0.462)
SH_ANN8	$5.20E-04$ (0.517)	-0.007 (-1.63)	0.010 (0.38)	-0.067^{***} (-2.83)	-0.004 (-0.436)	$2.91E-07$ (1.09)	0.040 (0.948)	0.044 (1.04)	-0.003 (-0.121)	-0.009 (-0.317)
SH_ANN9	$4.90E-04$ (0.521)	-0.007^{**} (-2.26)	0.009 (0.35)	-0.066^{***} (-2.90)	-0.005 (-0.523)	$3.22E-07$ (1.36)	0.039 (0.984)	0.044 (1.11)	-0.001 (-0.053)	-0.011 (-0.398)
SH_ANN10	$4.60E-04$ (0.495)	-0.006^{**} (-2.33)	0.011 (0.44)	-0.066^{***} (-2.71)	-0.005 (-0.552)	$2.84E-07$ (1.09)	0.037 (0.897)	0.045 (1.09)	$5.70E-04$ (0.019)	-0.008 (-0.261)

Values in parentheses denote z-statistics

***, **, * $\Rightarrow$ Significance at 1%, 5%, 10% level

Table 4.5: *Marginal effects for Layawan Watershed models*

	Variables						
Model	*AGE*	*AMOUNT*	*COLLEGE*	*GENDER*	*INC*	*IPAYEE*	*PAYSCH*
LW_Logit	0.002 (1.06)	-0.004^{***} (-12.48)	0.081 (1.61)	-0.049 (-1.04)	$5.00E-06^{*}$ (1.85)	-0.051 (-1.19)	-0.026 (-0.61)
LW_Probit	0.002 (1.09)	-0.009^{***} (-14.16)	0.081 (1.58)	-0.051 (-1.07)	$4.63E-06^{*}$ (1.90)	-0.048 (-1.12)	-0.029 (-0.68)
LW_KS	$-4.40E-04$ (1.09)	-0.016^{***} (14.16)	-0.013 (-0.20)	0.157 $(—)$	0.015^{*} (1.90)	-0.002 (1.12)	0.020 (0.68)
LW_ANN1	0.001^{*} (1.67)	-0.006^{***} (-3.93)	0.077^{*} (1.65)	-0.024 (-0.84)	$2.30E-06$ (1.09)	-0.045^{*} (-1.77)	-0.022 (-0.79)
LW_ANN2	0.002^{**} (2.19)	-0.006^{***} (-4.80)	0.101^{***} (2.92)	-0.030 (-0.86)	$2.73E-06^{*}$ (1.73)	-0.061^{*} (-1.79)	-0.017 (-0.55)
LW_ANN3	0.002 (1.29)	-0.006^{***} (-4.19)	0.058 (1.59)	-0.027 (-0.70)	$5.18E-06^{***}$ (2.87)	-0.047 (-1.64)	-0.016 (-0.35)
LW_ANN4	$8.70E-04$ (0.77)	-0.005^{***} (-4.14)	0.073^{**} (2.56)	-0.055 (-1.49)	$5.91E-06^{***}$ (3.38)	-0.073^{**} (-2.27)	-0.035 (-1.03)
LW_ANN5	$5.80E-04$ (0.55)	-0.006^{***} (-5.88)	0.086^{**} (2.43)	-0.017 (-0.54)	$5.31E-06^{***}$ (3.57)	-0.071^{**} (-2.43)	-0.026 (-1.05)
LW_ANN6	$5.00E-04$ (0.56)	-0.006^{***} (-4.81)	0.074^{**} (2.22)	-0.019 (-0.60)	$6.31E-06^{***}$ (3.18)	-0.059^{*} (-1.85)	-0.027 (-1.07)
LW_ANN7	$8.60E-04$ (0.88)	-0.006^{***} (-4.95)	0.073^{**} (2.11)	-0.035 (-1.16)	$3.44E-06^{**}$ (2.52)	-0.037 (-0.98)	-0.005 (-0.15)

Values in parentheses denote z-statistics

$^{***}, ^{**}, ^{*}$ $\Rightarrow$ Significance at 1%, 5%, 10% level

Table 4.6: *Median willingness-to-pay estimates*

Model	Average Median WTP	Standard Error
Smoky Hill Watershed Models		
SH_Logit	\$0.10	8.85
SH_Probit	\$0.09	4.58
SH_KS	\$0.06	0.09
SH_ANN1	\$1.35	0.95
SH_ANN2	\$1.75	1.94
SH_ANN3	\$2.73	3.48
SH_ANN4	\$3.22	4.10
SH_ANN5	\$3.25	3.52
SH_ANN6	\$3.39	3.69
SH_ANN7	\$3.25	3.48
SH_ANN8	\$2.98	2.64
SH_ANN9	\$2.91	3.17
SH_ANN10	\$2.89	3.03
Layawan Watershed Models		
LW_Logit	₱63.24***	8.85
LW_Probit	₱66.02***	4.58
LW_KS	₱25.15**	0.09
LW_ANN1	₱47.05***	0.95
LW_ANN2	₱53.16***	1.94
LW_ANN3	₱52.13***	3.48
LW_ANN4	₱56.60***	4.10
LW_ANN5	₱51.30***	3.52
LW_ANN6	₱51.46***	3.69
LW_ANN7	₱50.51***	3.48

Values in parentheses denote z-statistics

***, **, * $\Rightarrow$ Significance at 1%, 5%, 10% level

References

Adams, R. M., Rosenzweig, C., Peart, R., Ritchie, J. T., McCarl, B. A., et al. (1990). Global climate change and US agriculture. *Nature*, *345*(6272), 219.

Adriano, N., Wilson, D., & Joao, E. (2011). Contingent valuation of improvement in basic sanitation services: a case study in Brazil. *Econ. Environ. Stud.*, *11*(3), 213–234.

Adusumilli, N. C., Rister, M. E., Lacewell, R. D., et al. (2011). Estimation of irrigation water demand: A case study for the Texas High Plains. In *Selected paper presented at the saea annual meeting, february* (pp. 5–8).

Amemiya, T. (1981). Qualitative response models: A survey. *Journal of economic literature*, *19*(4), 1483–1536.

An, M. Y. (2000). A semiparametric distribution for willingness to pay and statistical inference with dichotomous choice contingent valuation data. *American Journal of Agricultural Economics*, *82*(3), 487–500.

Arbuckle, J. G., Morton, L. W., & Hobbs, J. (2013). Farmer beliefs and concerns about climate change and attitudes toward adaptation and mitigation: Evidence from Iowa. *Climatic Change*, *118*(3-4), 551–563.

Arbuckle, J. G., Prokopy, L. S., Haigh, T., Hobbs, J., Knoot, T., Knutson, C., ... others (2013). Climate change beliefs, concerns, and attitudes toward adaptation and mitigation among farmers in the Midwestern United States. *Climatic Change*, *117*(4), 943–950.

Arbuckle Jr, J. G., Morton, L. W., & Hobbs, J. (2015). Understanding farmer perspectives on climate change adaptation and mitigation: The roles of trust in sources of climate information, climate change beliefs, and perceived risk. *Environment and behavior*, *47*(2), 205–234.

Arnade, C., & Kelch, D. (2007). Estimation of area elasticities from a standard profit function. *American Journal of Agricultural Economics*, *89*(3), 727–737.

Arnold, B. C., & Press, S. J. (1989). Compatible conditional distributions. *Journal of the American Statistical Association*, *84*(405), 152–156.

Arouna, A., & Dabbert, S. (2012). Estimating rural households' willingness to pay for water supply improvements: A benin case study using a semi-nonparametric bivariate probit approach. *Water international*, *37*(3), 293–304.

Bailey, K. W., & Womack, A. W. (1985). Wheat acreage response: A regional econometric investigation. *Journal of Agricultural and Applied Economics*, *17*(2), 171.

Bastin, F., Cirillo, C., & Toint, P. L. (2010). Estimating nonparametric random utility models with an application to the value of time in heterogeneous populations. *Transportation science*, *44*(4), 537–549.

Bauer, P., Thorpe, A., & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, *525*(7567), 47–55.

Bazaraa, M. S., Sherali, H. D., & Shetty, C. M. (2013). *Nonlinear programming: Theory and algorithms.* John Wiley & Sons.

Beale, M. H., Hagan, M. T., & Demuth, H. B. (2017). Neural Network Toolbox user's guide. *MathWorks, Inc.*

Bellarby, J., Foereid, B., & Hastings, A. (2008). *Cool farming: Climate impacts of agriculture and mitigation potential.* Greenpeace.

Bergtold, J. S. (2004). *Advances in applied econometrics: Binary discrete choice models, artificial neural networks, and asymmetries in the FAST multistage demand system* (Unpublished doctoral dissertation). Virginia Polytechnic Institute and State University.

Bergtold, J. S., Spanos, A., & Onukwugha, E. (2010). Bernoulli regression models: Revisiting the specification of statistical models with binary dependent variables. *Journal of Choice Modelling*, *3*(2), 1–28.

Billingsley, P. (2008). *Probability and measure*. John Wiley & Sons.

Bjørner, T. B., & Leth-Petersen, S. (2007). A dynamic random effects multinomial logit model of household car ownership. *Nationaløkonomisk Tidsskrift*.

Bradshaw, B., Dolan, H., & Smit, B. (2004). Farm-level adaptation to climatic variability and change: Crop diversification in the Canadian prairies. *Climatic Change*, *67*(1), 119–141.

Breiman, L. (1996). Bagging predictors. *Machine learning*, *24*(2), 123–140.

Bridges, D., & Tenkorang, F. (2009). Agricultural commodities acreage value elasticity over timeimplications for the 2008 Farm Bill. In *American society of business and behavioral sciences annual conference, las vegas, nv.*

Burke, I. C., Yonker, C., Parton, W., Cole, C., Schimel, D., & Flach, K. (1989). Texture, climate, and cultivation effects on soil organic matter content in US grassland soils. *Soil Science Society of America Journal*, *53*(3), 800–805.

Calderon, M. M., Anit, K. P. A., Palao, L. K. M., & Lasco, R. D. (2012). Households willingness to pay for improved watershed services of the Layawan Watershed in Oroquieta City, Philippines. *Journal of Sustainable Development*, *6*(1), 1.

Chakir, R., & Le Gallo, J. (2013). Predicting land use allocation in France: A spatial panel data analysis. *Ecological Economics*, *92*, 114–125.

Chang, H.-H. (2012). Does the use of eco-labels affect income distribution and income inequality of aquaculture producers in Taiwan? *Ecological Economics*, *80*, 101–108.

Chatterjee, K. (2011). Modelling the dynamics of bus use in a changing travel environment using panel data. *Transportation*, *38*(3), 487–509.

Chavas, J.-P., & Holt, M. T. (1990). Acreage decisions under risk: The case of corn and soybeans. *American Journal of Agricultural Economics*, *72*(3), 529–538.

Chembezi, D. M., & Womack, A. W. (1992). Regional acreage response for US corn and wheat: The effects of government programs. *Journal of Agricultural and Applied Economics*, *24*(1), 187.

Cooper, J. C. (2002). Flexible functional form estimation of willingness to pay using dichotomous choice data. *Journal of Environmental Economics and management*, *43*(2), 267–279.

Cooper, J. C. (2003). A joint framework for analysis of agri-environmental payment programs. *American journal of agricultural economics*, *85*(4), 976–987.

Cooper, J. C., & Signorello, G. (2008). Farmer premiums for the voluntary adoption of conservation plans. *Journal of Environmental Planning and Management*, *51*(1), 1–14.

Cosslett, S. R. (1983). Distribution-free maximum likelihood estimator of the binary choice model. *Econometrica: Journal of the Econometric Society*, 765–782.

Creel, M., & Loomis, J. (1997). Semi-nonparametric distribution-free dichotomous choice contingent valuation. *Journal of Environmental Economics and Management*, *32*(3), 341–358.

Davidson, R., MacKinnon, J. G., et al. (1993). Estimation and inference in econometrics.

Diggs, D. M. (1991). Drought experience and perception of climatic change among Great Plains farmers. *Great Plains Research*, 114–132.

Fausett, L. V. (1994). *Fundamentals of neural networks*. Prentice-Hall.

Fine, T. L. (2006). *Feedforward neural network methodology*. Springer Science & Business Media.

Fok, Y. K., Jeon, S.-H., & Wilkins, R. (2013). Does part-time employment help or hinder single mothers movements into full-time employment? *Oxford Economic Papers*, *65*(2), 523–547.

Fosgerau, M. (2006). Investigating the distribution of the value of travel time savings. *Transportation Research Part B: Methodological*, *40*(8), 688–707.

Fosgerau, M., et al. (2005). Unit income elasticity of the value of travel time savings. In *8th nectar conference, las palmas gc june* (pp. 2–4).

Gabler, S., Laisney, F., & Lechner, M. (1993). Seminonparametric estimation of binary-choice models with an application to labor-force participation. *Journal of Business & Economic Statistics*, *11*(1), 61–80.

Gallant, A. R., & Nychka, D. W. (1987). Semi-nonparametric maximum likelihood estimation. *Econometrica: Journal of the Econometric Society*, 363–390.

Gong, X., Van Soest, A., & Villagomez, E. (2004). Mobility in the urban labor market: A panel data analysis for Mexico. *Economic Development and Cultural Change*, *53*(1), 1–36.

Greene, W. (2012). LIMDEP Version 10/NLOGIT Version 5. *Econometric Modeling Guide Plainview*.

Greene, W. H. (2012). *Econometric analysis 7th edition*. Prentice Hall.

Haden, V. R., Niles, M. T., Lubell, M., Perlman, J., & Jackson, L. E. (2012). Global and local concerns: What attitudes and beliefs motivate farmers to mitigate and adapt to climate change? *PloS one*, *7*(12), e52882.

Hanemann, W. M. (1984). Welfare evaluations in contingent valuation experiments with discrete responses. *American journal of agricultural economics*, *66*(3), 332–341.

Hanemann, W. M. (1991). Willingness to pay and willingness to accept: How much can they differ? *The American Economic Review*, *81*(3), 635–647.

Heckman, J. J. (1981). The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process.

Hendricks, N. P., Smith, A., & Sumner, D. A. (2014). Crop supply dynamics and the illusion of partial adjustment. *American Journal of Agricultural Economics*, aau024.

Hennessy, D. A. (2006). On monoculture and the structure of crop rotations. *American Journal of Agricultural Economics*, *88*(4), 900–914.

Hertel, T. W., Golub, A. A., Jones, A. D., O'Hare, M., Plevin, R. J., & Kammen, D. M. (2010). Effects of US maize ethanol on global land use and greenhouse gas emissions: Estimating market-mediated responses. *BioScience*, *60*(3), 223–231.

Hirschman, A. O. (1984). Against parsimony: Three easy ways of complicating some categories of economic discourse. *Bulletin of the American Academy of Arts and Sciences*, *37*(8), 11–28.

Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural networks*, *4*(2), 251–257.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, *2*(5), 359–366.

Hornik, K., Stinchcombe, M., & White, H. (1990). Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural networks*, *3*(5), 551–560.

Horowitz, J. L. (1993). 2 semiparametric and nonparametric estimation of quantal response models. *Handbook of statistics*, *11*, 45–72.

Horowitz, J. L. (2009). *Semiparametric and nonparametric methods in econometrics* (Vol. 12). Springer.

Howden, S. M., Soussana, J.-F., Tubiello, F. N., Chhetri, N., Dunlop, M., & Meinke, H. (2007). Adapting agriculture to climate change. *Proceedings of the national academy of sciences*, *104*(50), 19691–19696.

Huang, H., Khanna, M., et al. (2010). An econometric analysis of US crop yield and cropland acreage: Implications for the impact of climate change. In *Aaea annual meeting, denver, colorado* (pp. 25–27).

Isik, M., & Devadoss, S. (2006). An analysis of the impact of climate change on crop yields and yield variability. *Applied Economics*, *38*(7), 835–844.

Karlen, D. L., Berry, E. C., Colvin, T. S., & Kanwar, R. S. (1991). Twelve-year tillage and crop rotation effects on yields and soil chemical properties in northeast Iowa. *Communications in Soil Science & Plant Analysis*, *22*(19-20), 1985–2003.

Kastens, T. L., & Featherstone, A. M. (1996). Feedforward backpropagation neural networks in prediction of farmer risk preferences. *American Journal of Agricultural Economics*, 400–415.

Kay, R., & Little, S. (1987). Transformations of the explanatory variables in the logistic regression model for binary data. *Biometrika*, 495–501.

Klein, R. W., & Spady, R. H. (1993). An efficient semiparametric estimator for binary response models. *Econometrica: Journal of the Econometric Society*, 387–421.

Kmenta, J. (1986). *Elements of econometrics*. New York: Macrnilian.

Krinsky, I., & Robb, A. L. (1986). On approximating the statistical properties of elasticities. *The Review of Economics and Statistics*, 715–719.

Kuan, C.-M., & White, H. (1994). Artificial neural networks: An econometric perspective. *Econometric reviews*, *13*(1), 1–91.

Lal, R. (2004). Soil carbon sequestration impacts on global climate change and food security. *science*, *304*(5677), 1623–1627.

Lee, D. R., & Helmberger, P. G. (1985). Estimating supply response in the presence of farm programs. *American Journal of Agricultural Economics*, 193–203.

Leshno, M., Lin, V. Y., Pinkus, A., & Schocken, S. (1993). Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural networks*, *6*(6), 861–867.

Lichtenberg, E. (1989). Land quality, irrigation development, and cropping patterns in the Northern High Plains. *American Journal of Agricultural Economics*, *71*(1), 187–194.

Lin, B. B. (2011). Resilience in agriculture through crop diversification: Adaptive management for environmental change. *BioScience*, *61*(3), 183–193.

Lin, W., & Dismukes, R. (2007). Supply response under risk: Implications for countercyclical payments' production impact. *Review of Agricultural Economics*, 64–86.

Lobell, D. B., Burke, M. B., Tebaldi, C., Mastrandrea, M. D., Falcon, W. P., & Naylor, R. L. (2008). Prioritizing climate change adaptation needs for food security in 2030. *Science*, *319*(5863), 607–610.

Lobell, D. B., Schlenker, W., & Costa-Roberts, J. (2011). Climate trends and global crop production since 1980. *Science*, *333*(6042), 616–620.

Long, S. P., Ainsworth, E. A., Leakey, A. D., Nösberger, J., & Ort, D. R. (2006). Food for thought: Lower-than-expected crop yield stimulation with rising CO2 concentrations. *Science*, *312*(5782), 1918–1921.

Lubowski, R. N., Bucholtz, S., Claassen, R., Roberts, M. J., Cooper, J. C., Gueorguieva, A., & Johansson, R. (2006). Environmental effects of agricultural land-use change. *Economic research report*, *25*, 1–75.

Luenberger, D. G. (1969). *Optimization by vector space methods*. John Wiley & Sons.

Malone, L. A. (1986). A historical essay on the conservation provisions of the 1985 Farm Bill: Sodbusting, swampbusting, and the conservation reserve.

McIntosh, C. S., & Shideed, K. H. (1989). The effect of government programs on acreage response over time: The case of corn production in Iowa. *Western Journal of Agricultural Economics*, 38–44.

Mehrotra, K., Mohan, C. K., & Ranka, S. (1997). *Elements of artificial neural networks*. MIT press.

Mendelsohn, R., Nordhaus, W. D., & Shaw, D. (1994). The impact of global warming on agriculture: A Ricardian analysis. *The American economic review*, 753–771.

Miao, R., Khanna, M., & Huang, H. (2016). Responsiveness of crop yield and acreage to prices and climate. *American Journal of Agricultural Economics*, *98*(1), 191–211.

Morton, L. W., Hobbs, J., Arbuckle, J. G., & Loy, A. (2015). Upper Midwest climate variations: Farmer responses to excess water risks. *Journal of environmental quality*, *44*(3), 810–822.

Neal, R. M. (1996). Bayesian learning for neural networks.

Nguyen, D., & Widrow, B. (1990). Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights. In *Neural networks, 1990., 1990 ijcnn international joint conference on* (pp. 21–26).

Olesen, J. E., & Bindi, M. (2002). Consequences of climate change for European agricultural productivity, land use and policy. *European journal of agronomy*, *16*(4), 239–262.

Orazem, P. F., & Miranowski, J. A. (1994). A dynamic model of acreage allocation with general and crop-specific soil capital. *American Journal of Agricultural Economics*, *76*(3), 385–395.

Oregon Department of Environmental Quality. (2005, 04). Erosion and sediment control manual appendix b [Computer software manual]. Retrieved from http://www.deq.state.or.us/wq/stormwater/docs/escmanual/appxb.pdf

Parry, M. L., Rosenzweig, C., Iglesias, A., Livermore, M., & Fischer, G. (n.d.).

Plantinga, A. J., Mauldin, T., & Miller, D. J. (1999). An econometric analysis of the costs of sequestering carbon in forests. *American Journal of Agricultural Economics*, *81*(4), 812–824.

Plevin, R. J., O'Hare, M., Jones, A. D., Torn, M. S., & Gibbs, H. K. (2010). Greenhouse gas emissions from biofuels' indirect land use change are uncertain but may be greater than previously estimated. *Environmental Science & Technology*, *44*, 8015–8021.

Powers, D., & Xie, Y. (2008). *Statistical methods for categorical data analysis*. Emerald Group Publishing.

Principe, J. C., Euliano, N. R., & Lefebvre, W. C. (2000). *Neural and adaptive systems: Fundamentals through simulations* (Vol. 672). Wiley New York.

PRISM Climate Group. (2013, 06). Retrieved from http://prism.oregonstate.edu

Ramsey, J. B. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, 350–371.

Reynolds, C., Yitayew, M., Slack, D., Hutchinson, C., Huete, A., & Petersen, M. (2000). Estimating crop yields and production by integrating the FAO Crop Specific Water Balance model with real-time satellite data and ground-based ancillary data. *International Journal of Remote Sensing*, *21*(18), 3487–3508.

Reynolds, M. P., Quilligan, E., Aggarwal, P. K., Bansal, K. C., Cavalieri, A. J., Chapman, S. C., ... others (2016). An integrated approach to maintaining cereal productivity under climate change. *Global Food Security*, *8*, 9–18.

Ripley, B. D. (1994). Neural networks and related methods for classification. *Journal of the Royal Statistical Society. Series B (Methodological)*, 409–456.

Rosenzweig, C., Parry, M. L., et al. (1994). Potential impact of climate change on world food supply. *Nature*, *367*(6459), 133–138.

Rozeboom, K., Holman, J., Thompson, C., Shoup, D., Peterson, D., Duncant, S., & Olson, B. (2009). *No-till in Kansas: Rotations and cultural practices.* Kansas State University.

Schlenker, W., & Roberts, M. J. (2009). Nonlinear temperature effects indicate severe damages to US crop yields under climate change. *Proceedings of the National Academy of sciences*, *106*(37), 15594–15598.

Searchinger, T., Heimlich, R., Houghton, R. A., Dong, F., Elobeid, A., Fabiosa, J., ... Yu, T.-H. (2008). Use of US croplands for biofuels increases greenhouse gases through emissions from land-use change. *Science*, *319*(5867), 1238–1240.

Sen, A. K. (1977). Rational fools: A critique of the behavioral foundations of economic theory. *Philosophy & Public Affairs*, 317–344.

Seo, S. N., & Mendelsohn, R. (2008). An analysis of crop choice: Adapting to climate change in South American farms. *Ecological economics*, *67*(1), 109–116.

Small, C. G., & McLeish, D. L. (2011). *Hilbert space methods in probability and statistical inference* (Vol. 920). John Wiley & Sons.

Smit, B., & Skinner, M. W. (2002). Adaptation options in agriculture to climate change: A typology. *Mitigation and adaptation strategies for global change*, *7*(1), 85–114.

Smith, P., Martino, D., Cai, Z., Gwary, D., Janzen, H., Kumar, P., ... others (2008). Greenhouse gas mitigation in agriculture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1492), 789–813.

Spanos, A. (1999). *Probability theory and statistical inference: Econometric modeling with observational data.* Cambridge University Press.

Staggenborg, S. A., Dhuyvetter, K. C., & Gordon, W. (2008). Grain sorghum and corn comparisons: Yield, economic, and environmental responses. *Agronomy journal*, *100*(6), 1600–1604.

Stone, L. R., & Schlegel, A. J. (2006). Crop water use in limited-irrigation environments. In *Proceedings of the 2006 Central Plains Irrigation Conference, Colby, KS* (pp. 21–23).

Sugden, R. (1999). Alternatives to the neoclassical theory of choice. *Valuing Environmental Preferences*, 152–180.

Theil, H. (1969). A multinomial extension of the linear logit model. *International Economic Review*, *10*, 251–259.

Train, K. (2003). *Discrete choice methods with simulation.* Cambridge university press.

United States Department of Agriculture, National Agricultural Statistics Service. (2016, 04). Cropland Data Layer [Computer software manual]. Retrieved from https://nassgeodata.gmu.edu/CropScape/

United States Department of Agriculture, National Agricultural Statistics Service. (2017, 02). QuickStats [Computer software manual]. Retrieved from https://quickstats.nass.usda.gov/

United States Department of Agriculture, National Resources Conservation Service. (2016, 11). Gridded Soil Survey Geographic [Computer software manual]. Retrieved from https://datagateway.nrcs.usda.gov

United States Department of Agriculture, Natural Resources Conservation Service. (2014). Ssurgo 2.3.2 table column descriptions.

United States Department of Agriculture, Natural Resources Conservation Service. (2017, 03). Updated T and K factors [Computer software manual]. Retrieved from https://www.nrcs.usda.gov/wps/PA_NRCSConsumption/download?cid=stelprdb1262856&ext=pdf

United States Geological Survey. (2017, 05). *National Water Information System.* Retrieved from https://waterdata.usgs.gov/nwis

West, P. M., Brockett, P. L., & Golden, L. L. (1997). A comparative analysis of neural networks and statistical methods for predicting consumer choice. *Marketing Science*, *16*(4), 370–391.

White, H. (1989). Some asymptotic results for learning in single hidden-layer feedforward network models. *Journal of the American Statistical Association*, *84*(408), 1003–1013.

Wooldridge, J. M. (2005). Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of applied econometrics*, *20*(1), 39–54.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Wu, J., & Adams, R. M. (2002). Micro versus macro acreage response models: Does site-specific information matter? *Journal of Agricultural and resource Economics*, 40–60.

Wu, J., & Adams, R. M. (2003). Production risks, acreage decisions, and implications for revenue insurance programs. In *Risk management and the environment: Agriculture in perspective* (pp. 161–180). Springer.

Wu, J., & Brorsen, B. W. (1995). The impact of government programs and land characteristics on cropping patterns. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie*, *43*(1), 87–104.

Wu, J., Mapp, H. P., & Bernardo, D. J. (1996). Integrating economic and physical models for analyzing water quality impacts of agricultural policies in the High Plains. *Review of Agricultural Economics*, 353–372.

Wu, J., & Segerson, K. (1995). The impact of policies and land characteristics on potential groundwater pollution in Wisconsin. *American Journal of Agricultural Economics*, *77*(4), 1033–1047.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American statistical Association*, *57*(298), 348–368.