An artificial neural network approach to acreage-share modeling

Steven M. Ramsey

Assistant Professor

New Mexico State University, Dept. of Agricultural Economics & Agricultural Business

smramsey@nmsu.edu

Jason S. Bergtold

Professor

Kansas State University, Department of Agricultural Economics

bergtold@ksu.edu

Jessica L. Heier-Stamm

Assisstant Professor

Kansas State University, Department of Industrial and Manufacturing Systems Engineering

jlhs@ksu.edu

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

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

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 provided in section 2. Section 3 outlines the empirical methods for both approaches as well as elasticity calculations (one basis of comparison) under each. Section 4 provides a description of the empirical application and associated data. In addition to estimated-elasticity comparisons,

---

[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.

the two approaches are compared with respect to model fit. The results of these comparisons are presented in section 5. Concluding remarks and discussion are provided in section 6.

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:

$$\Pi_j > \Pi_k \quad \text{for all } k \neq j, \tag{2}$$

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 2 can be rewritten as:

$$\pi_j + \varepsilon_j > \pi_k + \varepsilon_j \quad \text{for all } k \neq j. \tag{3}$$

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

$$\pi_j = g\left(\mathbf{x}_j; \boldsymbol{\beta}_j\right), \tag{4}$$

where $\mathbf{x}$ is a vector of explanatory variables and $\boldsymbol{\beta}$ is a vector of parameters. Then, using equation 4, condition 3 becomes:

$$g\left(\mathbf{x}_j; \boldsymbol{\beta}_j\right) + \varepsilon_j > g\left(\mathbf{x}_k; \boldsymbol{\beta}_k\right) + \varepsilon_k \quad \text{for all } k \neq j. \tag{5}$$

Decomposing the farmer's problem as in condition 5 allows it to be viewed from a probabilistic perspective. That is, the probability that crop j is planted, P_j, is:

$$P_j = P\left(g\left(\mathbf{x}_j; \boldsymbol{\beta}_j\right) + \varepsilon_j > g\left(\mathbf{x}_k; \boldsymbol{\beta}_k\right) + \varepsilon_k \quad \text{for all } k \neq j\right) \tag{6}$$

or

$$P_j = P\left(\varepsilon_k - \varepsilon_j < g\left(\mathbf{x}_j; \boldsymbol{\beta}_j\right) - g\left(\mathbf{x}_k; \boldsymbol{\beta}_k\right) \quad \text{for all } k \neq j\right). \tag{7}$$

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

$$s_j = P_j = \frac{\exp\left(g\left(\mathbf{x}_j; \boldsymbol{\beta}_j\right)\right)}{\sum_{j=0}^{J} \exp\left(g\left(\mathbf{x}_j; \boldsymbol{\beta}_j\right)\right)}. \tag{8}$$

The identity given in equation 8 provides the basis for the empirical procedures outlined in section 3.

3 Empirical Models

Two approaches are used to estimate the share equations given in equation 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).

As noted by Greene (2012), an indeterminacy arises in the model when $g\left(\cdot\right)$ is taken to be the linear function $g\left(\mathbf{x}_j; \boldsymbol{\beta}_j\right) = \mathbf{x}_j' \boldsymbol{\beta}_j$. 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 $\mathbf{q}$ (Greene, 2012). This problem can be resolved by setting all parameters for the 0th crop equal to zero; that is, setting $\boldsymbol{\beta}_0 = \mathbf{0}$ (Greene, 2012). Then, as long as $g\left(\mathbf{x}_0; \boldsymbol{\beta}_0\right) = 0$, equation 8 becomes

$$s_j = P_j = \frac{\exp\left(g\left(\mathbf{x}_j; \boldsymbol{\beta}_j\right)\right)}{1 + \sum_{j=1}^{J} \exp\left(g\left(\mathbf{x}_j; \boldsymbol{\beta}_j\right)\right)} \quad \text{for } j = 1, \ldots, J \tag{9}$$

and

$$s_0 = P_0 = \frac{1}{1 + \sum_{j=1}^{J} \exp\left(g\left(\mathbf{x}_j; \boldsymbol{\beta}_j\right)\right)} \quad \text{for } j = 0. \tag{10}$$

It will be shown in subsections 3.1 and 3.2 that $g\left(\mathbf{x}_0; \boldsymbol{\beta}_0\right) = 0$ if $\boldsymbol{\beta}_0 = \mathbf{0}$ for both the SUR-HEAR and ANN approaches.

Noting that equation 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 9 by s_0 in equation 10:

$$\frac{s_j}{s_0} = \exp\left(g\left(\mathbf{x}_j; \boldsymbol{\beta}_j\right)\right) \quad \text{for } j = 1, \ldots, J. \tag{11}$$

Taking the natural logarithm of equation 11 leaves the estimable system

$$\ln\left(\frac{s_j}{s_0}\right) = g\left(\mathbf{x}_j; \boldsymbol{\beta}_j\right) \text{ for } j = 1, \ldots, J. \tag{12}$$

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 3.2 below.

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 (Greene, 2012). Letting $i = 1, \ldots, N$ denote regions and $t = 1, \ldots, T$ denote time periods, this assumption implies that

$$E\left(\boldsymbol{\varepsilon}_j \boldsymbol{\varepsilon}_j' \mid \mathbf{X}_1, \ldots, \mathbf{X}_J\right) = \sigma_{jj} \mathbf{I}_{NT} \tag{13}$$

where $\boldsymbol{\varepsilon}_j = [\varepsilon_{1,1,j} \ldots \varepsilon_{1,T,j} \ldots \varepsilon_{N,1,j} \ldots \varepsilon_{N,T,j}]'$ and NT denotes the total number of observations for each crop-share $j = 1, \ldots, J$. Wu and Brorsen (1995) suggest that condition 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 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 (Greene, 2012):

$$E\left(\varepsilon_{i,t,j}\varepsilon_{m,s,k} \mid \mathbf{X}_1, \ldots, \mathbf{X}_J\right) = \begin{cases} \sigma_{j,k} & \text{for } t = s, i = m \\ 0 & \text{for } t \neq s \text{ or } i \neq m \end{cases}. \tag{14}$$

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

$$s_j = P_j = \frac{\exp\left(\mathbf{x}_j' \boldsymbol{\beta}_j\right)}{1 + \sum_{j=1}^{J} \exp\left(\mathbf{x}_j' \boldsymbol{\beta}_j\right)} \tag{15}$$

and the estimated equations become

$$\ln\left(\frac{s_j}{s_0}\right) = \mathbf{x}_j' \boldsymbol{\beta}_j. \tag{16}$$

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 9 and 10.

The SUR-HEAR approach starts by estimating each equation given in equation 16 via ordinary least squares (OLS). Using the regression residuals, $\tilde{\varepsilon}_{i,t,j}$, the autocorrelation coefficient, $\rho_{i,j}$, is estimated as

$$\tilde{\rho}_{i,j} = \frac{\sum_{t=2}^{T} \tilde{\varepsilon}_{i,t,j} \tilde{\varepsilon}_{i,t-1,j}}{\sum_{t=1}^{T} \tilde{\varepsilon}_{i,t,j}^2}. \tag{17}$$

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

$$y_{i,1,j}^* = y_{i,1,j} \sqrt{1 - \tilde{\rho}_{i,j}^2}, \tag{18}$$

$$y_{i,t,j}^* = y_{i,t,j} - \tilde{\rho}_{i,j} y_{i,t-1,j} \text{ for } t = 2, \ldots, T, \tag{19}$$

$$x_{i,1,j}^* = x_{i,1,j} \sqrt{1 - \tilde{\rho}_{i,j}^2}, \tag{20}$$

and

$$x_{i,t,j}^* = x_{i,t,j} - \tilde{\rho}_{i,j} x_{i,t-1,j} \text{ for } t = 2, \ldots, T, \tag{21}$$

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 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:

$$\hat{\sigma}_{i,j} = \sum_{t=1}^{T} \frac{\left(\tilde{\varepsilon}_{i,t,j}^*\right)^2}{T}. \tag{22}$$

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

$$y_{i,t,j}^{**} = \frac{y_{i,t,j}^*}{\hat{\sigma}_{i,j}} \tag{23}$$

and

$$x_{i,t,j}^{**} = \frac{x_{i,t,j}^*}{\hat{\sigma}_{i,j}}. \tag{24}$$

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

$$y_{i,t,j}^{**} = \mathbf{x}_{i,t,j}^{**\prime} \boldsymbol{\beta}_j + \varepsilon_{i,t,j}^{**}. \tag{25}$$

6

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

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), but a simple identity function may be used as well. An intercept term can also be added to yield (Fausett, 1994):

$$net_i = \beta_0 + \sum_{k=1}^{K} \beta_k x_{i,k} \tag{26}$$

and

$$y_i = \mathcal{F}(net_i) = \mathcal{F}\left(\beta_0 + \sum_{k=1}^{K} \beta_k x_{i,k}\right). \tag{27}$$

If $\mathcal{F}$ is the identity function, equation 27 is simply

$$y_i = \mathcal{F}(net_i) = \beta_0 + \sum_{k=1}^{K} \beta_k x_{i,k}. \tag{28}$$

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:

$$y_{i,h} = \mathcal{F}_1 \left(net_{i,h} \right) \text{ for } h = 1, \dots, H, \tag{29}$$

where

$$net_{i,h} = \beta_{0,h} + \sum_{k=1}^{K} \beta_{k,h} x_{i,k} \tag{30}$$

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

$$y_i = \mathcal{F}_2 \left(net_i \right), \tag{31}$$

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

$$net_i = \delta_0 + \sum_{h=1}^{H} \delta_h \mathcal{F}_1 \left(\beta_{0,h} + \sum_{k=1}^{K} \beta_{k,h} x_{i,k} \right). \tag{32}$$

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 $j = 1, \dots, J$ output neurons, where the weights are unique to each hidden neuron-output neuron pair. For this architecture, outputs are given by

$$y_{i,j} = \mathcal{F}_2 \left(net_{i,j} \right) \tag{33}$$

where

$$net_{i,j} = \delta_{0,j} + \sum_{h=1}^{H} \delta_{h,j} \mathcal{F}_1 \left(\beta_{0,h} + \sum_{k=1}^{K} \beta_{k,h} x_{i,k} \right). \tag{34}$$

If $\mathcal{F}_2$ is chosen to be a purely linear function, then $y_{i,j}$ is simply equal to equation 34, 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 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 34. Under this framework, the share of crop $j = 1, \dots, J$

8

is given by

$$s_j = P_j = \frac{\exp\left(\delta_{0,j} + \sum_{h=1}^{H} \delta_{h,j} \mathcal{F}_1\left(\beta_{0,h} + \sum_{k=1}^{K} \beta_{k,h} x_{i,k}\right)\right)}{\sum_{j=0}^{J} \exp\left(\delta_{0,j} + \sum_{h=1}^{H} \delta_{h,j} \mathcal{F}_1\left(\beta_{0,h} + \sum_{k=1}^{K} \beta_{k,h} x_{i,k}\right)\right)} \tag{35}$$

or more compactly by

$$s_j = P_j = \frac{\exp\left(\delta_{0,j} + \mathbf{y}_H' \boldsymbol{\delta}_j\right)}{\sum_{j=0}^{J} \exp\left(\delta_{0,j} + \mathbf{y}_H' \boldsymbol{\delta}_j\right)} \tag{36}$$

where $\mathbf{y}_H' = [y_{i,1} \ y_{i,2} \cdots y_{i,H}]$ and $\boldsymbol{\delta}_j = [\delta_{1,j} \ \delta_{2,j} \cdots \delta_{h,j}]'$. Note that the same indeterminancy arises for this specification with respect to the δ parameters. This indeterminancy is removed by setting $\boldsymbol{\delta}_0 = \mathbf{0}$, which yields $g(\mathbf{x}_0; \boldsymbol{\beta}, \boldsymbol{\delta}_0) = 0$, as required. By doing so, the share equations become

$$s_j = P_j = \frac{\exp\left(\delta_{0,j} + \mathbf{y}_H' \boldsymbol{\delta}_j\right)}{1 + \sum_{j=1}^{J} \exp\left(\delta_{0,j} + \mathbf{y}_H' \boldsymbol{\delta}_j\right)} \quad \text{for } j = 1, \ldots, J \tag{37}$$

and

$$s_j = P_j = \frac{1}{1 + \sum_{j=1}^{J} \exp\left(\delta_{0,j} + \mathbf{y}_H' \boldsymbol{\delta}_j\right)} \quad \text{for } j = 0. \tag{38}$$

The estimated equations from expression 12 then become

$$\ln\left(\frac{s_j}{s_0}\right) = \delta_{0,j} + \sum_{h=1}^{H} \delta_{h,j} \mathcal{F}_1\left(\beta_{0,h} + \sum_{k=1}^{K} \beta_{k,h} x_{i,k}\right). \tag{39}$$

The structure of an ANN given by equation 34 is represented in figure 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

$$MSE = \frac{1}{J} \frac{1}{T} \frac{1}{N} \sum_{j=1}^{J} \sum_{t=1}^{T} \sum_{i=1}^{N} \left(\hat{y}_{i,t,j} - y_{i,t,j}\right)^2, \tag{40}$$

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 Elasticities

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

$$\eta_{i,t,j,k} = \frac{\partial A_{i,t,j}}{\partial x_{i,t,j,k}} \frac{x_{i,t,j,k}}{A_{i,t,j}} \tag{41}$$

where $A_{i,t,j}$ is the total acres in region i allocated to crop j at time t. Using land-use shares, equation 41 can be rewritten as

$$\eta_{i,t,j,k} = \frac{\partial s_{i,t,j} \bar{A}_i}{\partial x_{i,t,j,k}} \frac{x_{i,t,j,k}}{s_{i,t,j} \bar{A}_i} \tag{42}$$

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

$$\eta_{i,t,j,k} = \frac{\partial P_{i,t,j} \bar{A}_i}{\partial x_{i,t,j,k}} \frac{x_{i,t,j,k}}{P_{i,t,j} \bar{A}_i} \tag{43}$$

or

$$\eta_{i,t,j,k} = \frac{\partial P_{i,t,j}}{\partial x_{i,t,j,k}} \frac{x_{i,t,j,k}}{P_{i,t,j}}, \tag{44}$$

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

$$ME_{i,t,j,k} = P_{i,t,j} \left(\beta_{j,k} - \sum_{j=1}^{J} \beta_{j,k} P_{i,t,j} \right), \tag{45}$$

while for the ANN approach it is given by

$$\begin{aligned} ME_{i,t,j,k} = \quad & P_{i,t,j} \left(\sum_{h=1}^{H} \delta_{h,j} \mathcal{F}_1 \left(net_{i,h}\right) \left[1 - \mathcal{F}_1 \left(net_{i,h}\right)\right] \beta_{k,h} - \right. \\ & \left. \sum_{j=1}^{J} \sum_{h=1}^{H} \delta_{h,j} \mathcal{F}_1 \left(net_{i,h}\right) \left[1 - \mathcal{F}_1 \left(net_{i,h}\right)\right] \beta_{k,h} P_{i,t,j} \right), \end{aligned} \tag{46}$$

11

where $net_{i,h}$ is as given in equation 30. Using equations 43-46, the elasticities become

$$\eta_{i,t,j,k} = x_{i,t,j,k} \left(\beta_{j,k} - \sum_{j=1}^{J} \beta_{j,k} P_{i,t,j} \right) \tag{47}$$

for the SUR-HEAR model and

$$\eta_{i,t,j,k} = x_{i,t,j,k} \left(\sum_{h=1}^{H} \delta_{h,j} \mathcal{F}_1 (net_{i,h}) [1 - \mathcal{F}_1 (net_{i,h})] \beta_{k,h} - \sum_{j=1}^{J} \sum_{h=1}^{H} \delta_{h,j} \mathcal{F}_1 (net_{i,h}) [1 - \mathcal{F}_1 (net_{i,h})] \beta_{k,h} P_{i,t,j} \right) \tag{48}$$

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

$$\bar{A}_r = \sum_{i=1}^{N} \omega_{i,r} \bar{A}_i, \tag{49}$$

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

$$A_{r,t,j} = \sum_{i=1}^{N} \omega_{i,r} s_{i,t,j} \bar{A}_i \tag{50}$$

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

$$A_{r,t,j} = \sum_{i=1}^{N} \omega_{i,r} P_{i,t,j} \bar{A}_i. \tag{51}$$

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

$$ME_{r,t,j,k} = \frac{\partial A_{r,t,j}}{\partial \mathbf{x}_{r,t,k}} = \sum_{i=1}^{N} \frac{\partial \omega_{i,r} P_{i,t,j} \bar{A}_i}{\partial x_{i,t,j,k}} \tag{52}$$

or

$$ME_{r,t,j,k} = \sum_{i=1}^{N} \omega_{i,r} \bar{A}_i \frac{\partial P_{i,t,j}}{\partial x_{i,t,j,k}}. \tag{53}$$

The elasticity for the aggregated region can then be calculated as

$$\eta_{r,t,j,k} = \left[\frac{\partial \bar{A}_{r,t,j}}{\partial \mathbf{x}_{r,t,k}} \right]' \frac{\mathbf{x}_{r,t,k}}{\bar{A}_{r,t,j}} = \frac{1}{\bar{A}_{r,t,j}} \sum_{i=1}^{N} \omega_{i,r} \bar{A}_i \frac{\partial P_{i,t,j}}{\partial x_{i,t,j,k}} x_{i,t,j,k}, \tag{54}$$

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 54 becomes

$$\eta_{r,t,j,k} = \frac{1}{\bar{A}_{r,t,j}} \sum_{i=1}^{N} \omega_{i,r} \bar{A}_i P_{i,t,j} \left(\beta_{j,k} - \sum_{j=1}^{J} \beta_{j,k} P_{i,t,j} \right) x_{i,t,j,k} \tag{55}$$

and for the ANN approach becomes

$$\eta_{r,t,j,k} = \frac{1}{\bar{A}_{r,t,j}} \sum_{i=1}^{N} \omega_{i,r} \bar{A}_i P_{i,t,j} \left(\sum_{h=1}^{H} \delta_{h,j} \mathcal{F}_1 \left(net_h \right) \left[1 - \mathcal{F}_1 \left(neth_h \right) \right] \beta_{k,h} - \sum_{j=1}^{J} \sum_{h=1}^{H} \delta_{h,j} \mathcal{F}_1 \left(net_h \right) \left[1 - \mathcal{F}_1 \left(net_h \right) \right] \delta_{h,j} P_{i,t,j} \right) x_{i,t,j,k}. \tag{56}$$

An alternative, more compact way of writing equations 55 and 56 is

$$\eta_{r,t,j,k} = \sum_{i=1}^{N} \gamma_{i,r,j} \eta_{i,t,j,k} \tag{57}$$

where

$$\gamma_{i,r} = \frac{\omega_{i,r} \bar{A}_i P_{i,t,j}}{\bar{A}_{r,t,j}} \tag{58}$$

and $\eta_{i,t,j,k}$ is the sub-region elasticity given in either equation 47 or 48. 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 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).

3.4   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:

$$MSE_j = \frac{1}{N} \sum_{i=1}^{N} \left(s_{i,t,j} - \hat{s}_{i,j,t} \right)^2, \tag{59}$$

where $N = 2,342$. To get an average MSE for crop j across all years, equation 59 becomes:

$$MSE_j = \frac{1}{N \times T} \sum_{t=1}^{T} \sum_{i=1}^{N} \left(s_{i,t,j} - \hat{s}_{i,j,t} \right)^2, \tag{60}$$

where $T = 8$. 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:

$$MSE_j = \frac{1}{1000 \times N} \sum_{m=1}^{1000} \sum_{i=1}^{N} \left(s_{i,j,t,m} - \hat{s}_{i,j,t,m} \right)^2. \tag{61}$$

Similar adjustments can be made to equation 61 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 use:

$$MSE = \frac{1}{N \times T \times (J+1)} \sum_{j=0}^{4} \sum_{t=1}^{T} \sum_{i=1}^{N} \left(s_{i,t,j} - \hat{s}_{i,j,t} \right)^2 \tag{62}$$

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

$$MSE = \frac{1}{1,000 \times N \times T \times (J+1)} \sum_{m=1}^{1000} \sum_{j=0}^{4} \sum_{t=1}^{T} \sum_{i=1}^{N} \left(s_{i,t,j,m} - \hat{s}_{i,j,t,m} \right)^2 \tag{63}$$

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 51 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

$$D_{t,j} = \frac{\hat{A}_{r,t,j}}{A_{r,t,j}} - 1. \tag{64}$$

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.

4 Data

Kansas land-use data is used to compare the two approaches outlined in section 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 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} (90.1 km^2), or roughly 22,449 acres (9,084 ha).

A number of explanatory variables are included in the empirical models: 18 (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 4.1 to 4.4 below; summary statistics are found in tables 2 to 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. Due to the use of lagged-dependent variables, there are 18,736 total observations for each approach. An example of CDL imagery is provided in figure 3 and average shares (across township units) by year in table 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).

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, Smith, and Sumner (2014) and is given by

$$E\left(p_{t,j}\right) = FP_{t,j} + E\left(B_{t,j}\right). \tag{65}$$

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 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 for summary data on price variables.

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 (%Sand = 1 − %Clay − %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 (694 tonnes/ha/year) for a wide, barren field (Oregon Department of Environmental Quality, 2005).

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.

5 Results

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 Greene (2012, p. 962). The tests failed to reject the null hypotheis of no autocorrelation at the

10% level. A null hypothesis of no heteroskedasticity was rejected at the 1% level for each crop based on the Lagrange-multiplier test from 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 Greene (2012, p. 338). Test statistics are presented in table 4. Based on these results, the original data was corrected for hetroskedasticty but not autocorrelation.

To motivate the use of the ANNs, the Ramsey (1969) RESET test was used as described in 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 5, 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 (Greene, 2012). Due to their approximation capabilities, ANNs lessen the need for researchers to isolate the correct specification.

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, which will be referred to as ANN41. The average MSE across the 1,000 data partitions for ANN41 was 0.6736[2]. ANN performance across the number of hidden-layer neurons is presented in figure 5. The figure shows little variation in performance across networks for a number hidden nodes ranging from about seven to 43.

5.3 Model Fit

Model fit favors ANN41 for the measures examined here: MSE — with respect to predicted land-use shares — and deviations between the predicted and actual aggregate acres for the state. MSE results are presented in table 6 and figure 6. The values in these tables and figures were calculated using the methods outlined in equations 59 to 63. Across all crops and years, the SUR-HEAR MSE was 0.008. Averaged over the 1,000

---

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

runs, the ANN41 MSE across all crops and years was slightly lower at 0.005. The lowest MSE for both approaches was 0.002, seen in multiple years for sorghum. The largest SUR-HEAR MSE was 0.045 and for ANN41 was 0.01, both for the 2013 "other" estimate. For both approaches, when averaged across years, the smallest MSEs were for sorghum (0.003 for SUR-HEAR; 0.002 for ANN41) while the largest were "other" for the SUR-HEAR (0.012) and wheat for ANN41 (0.006).

Residual boxplots (figure 7) also indicate a slight advantage for the ANN approach. Figure 7 depicts township-level boxplots for each crop-model combination, grouped by actual shares. Residuals were calculated as $s_{i,t,j} - \hat{s}_{i,t,j}$ and so positive residuals indicate an under-estimation and negative residuals an over-estimateion. ANN residuals were obtained from the best-fitting[3] network (BANN) across the 1,000 ANN41 runs. A few reasons are offered for the use of the BANN. First, to aid with interpretation: the number of outliers would likely increase if all 1,000 ANN41 models were 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 5), the remaining networks were likely to generate similar results. The SUR-HEAR and BANN approaches produce similar results for the sorghum, soybeans, and wheat esimates. For soybeans, the 25th and 75th percentiles are tightly grouped around zero for townships with actual shares of roughly 0 to 60%, with slight upward movements thereafter. A similar pattern is seen with wheat residuals, though with a less pronounced upward movements at higher actual shares. There is a more pronounced upward movement in sorghum residuals for both models, starting around actual shares of between 20 and 30%. For the two remaining crops, corn and the "other" category, the SUR-HEAR approach exhibits more prounounced upward movements in the median values and the 25th and 75th percentiles. Additionally, the increase in interquartile range for the 25th and 75th percentiles at higher actual shares is larger for the SUR-HEAR approach, though a similar pattern is seen in the BANN residuals. Thus, though both approaches show some inability to predict higher levels of crop shares, the BANN shows some capacity limit this, relative to the SUR-HEAR approach, for corn and "other".

Effects of prediction performance at the individual township level subsequently impact predictions of state-level acreages. This can be seen in figure 8 and table 7, which show the percent deviations of predicted acreages from actual acreages for the SUR-HEAR, the BANN, and the worst[4] ANN (WANN). 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: in 11 of the 40 crop-year combinations, SUR-HEAR estimated acreages were +/-10% or more from the actual acreages, going as high as a 49% overestimation of soybean acres in 2013. Across all years, SUR-HEAR accuracy was best for wheat estimates, for which

---

[3]Based on the lowest average MSE across land uses.
[4]Based on the largest average MSE across land uses.

the mean absolute deviation (MAD) was about 3%. Wheat was also the only crop for which the SUR-HEAR estimated acreage did not over- or under-estimate the actual acreage by 10% in an year. The worst SUR-HEAR MAD was for the soybeans at about 15%. Generally, the BANN produced accurate acreage estimates, with the largest deviation being an 8% underestimate of corn in 2012. The BANN performed best with respect to wheat, with a MAD of about 1%, and generally was poorest for corn, with a MAD of roughly 4%. The largest ANN deviations were -8%, seen for 2011 sorghum (WANN) and 2012 corn (BANN). Both approaches tended to over-estimate the "other" share and under-estimate the corn share, with the exception of 2010 corn. Generally, despite MADs being smaller than the SUR-HEAR counterparts, the BANN and WANN more consistently over-estimated wheat and soybeans and under-estimated sorghum. 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.

5.4   Acreage Response

Annual acreage-response elasticities were calculated using the methods in section 3.3. Standard errors for the ANN41 results were obtained via the "bagging" approach described in section 3.2. Standard errors for the SUR-HEAR elasticities were estimated by bootstrapping the estimates across 1,000 runs. See section 3.4 for details. Bootstrapped/bagged elasticities, averaged across years and township units are shown in table 8.

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 60 elasticities in table 8 associated with the SUR-HEAR model, only three were 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 the lagged shares, the corn equation, for which the hypothesis of a linear index function was not rejected, was the only point of consistency, in terms of sign and statistical significance, between the SUR-HEAR and ANN approaches.

Across both approaches, own-price elasticities mostly held to economic theory: increased expected prices led to increased acres. The one exception was $PSOY$ in the SUR-HEAR model, where the estimated elasticity was -0.26. 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 $PWHEAT$ at 0.06, which was still larger than the ANN41 estimate

of 0.02. Corn and sorghum own-price elasticities from the SUR-HEAR were much larger, at 0.69 and 1.42 respectively. With ANN41, these two elasticities were 0.06 and 0.04, respectively. Though they are in the low range of own-price elasticities, the results from ANN41 appear to be more in line with those from past studies (see table 1). All cross-price elasticities from the SUR-HEAR were significant at the 1% level except for the impact of $PCORN$ on wheat. Only four (of 12) of the cross-price elasticities were significant using ANN41: $PSORGHUM$ and $PWHEAT$ on corn acres, $PCORN$ on sorghum, and $PWHEAT$ on soybeans. Elasticities associated with $PDIESEL$ were negative and significant for sorghum and wheat and positive and significant for corn and soybeans with the SUR-HEAR approach. $PLABOR$ had a positive and significant impact on corn, sorghum and wheat and a significant negative impact on soybeans in the SUR-HEAR model. All $PDIESEL$ and $PLABOR$ elasticities were insignificant with ANN41.

All but three elasticities associated with weather variables were statistically significant 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 ANN41, AVG_{CSS} had a negative and significant effect on sorghum acres and a positive and significant impact on soybean acres. ANN41 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 soybean acres and a significant negative impact on wheat acres with the SUR-HEAR model. The impact on soybean acres may reflects shifts from corn to one of the other two crops during wet years as corn is generally planted earlier. ANN41 estimated a positive and significant impact on corn, sorghum, and wheat for $PREC_{CSS}$. 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 for the SUR-HEAR approach except for the effect of $TMIN_{CSS}$ on soybeans. Only two significant temperature-based elasticities were found using ANN41: soybean acres are negatively affected by $TMAX_W$ and positively by $TMIN_{CSS}$.

Elasticity differences are also seen with respect to spatial distributions. The difference between SUR-HEAR- and ANN-estimated township-level elasticities are presented in figures 9 to 12. These values were calculated as the average township elasticity across years for the SUR-HEAR model minus the same measure from ANN41, averaged across years and the 1,000 runs. For almost all township units, the SUR-HEAR approach estimated larger own-price elasticities for corn and sorghum (figures 9 and 10). For corn, there is an east-west gradient in the differences, which generally increase moving from the western border to the

21

eastern border. This pattern is not as evident in the own-price sorghum elasticities, though spatial patterns can be seen. Differences tended to be larger in each of the four state corners, and smaller in north-central Kansas. No negative differences are seen in the sorghum map, and just five are seen in the corn map, located in southwest Kansas. In contrast, there are no positive values seen in the soybean map (figure 11). Differences in the soybean own-price elasticities were increasingly negative in the central and south-central portions of Kansas and decreasingly negative in eastern and western Kansas. The wheat own-price elasticities exhibited a little more parity between greater SUR-HEAR estimates and greater ANN41 estimates (figure 12). Large blocks of negative differences are seen in central and southwest Kansas; the largest positive differences are found in the northeast corner.

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 a misspecification of the sorghum, soybeans, and wheat equations.

Empirical results indicated some differences between the two approaches. The ANN approach shows a slight advantage in estimating observed shares (and thus actual acreages) compared with the SUR-HEAR approach. 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, ANN estimation procedures offer some protection against this with stopping criteria that use cross-validation. The ANN approach also produced elasticity estimates which were, for the most part, statistically insignificant and of a much smaller magnitude. Own- and cross-price

acreage elasticities largely held to expectations and economic theory across both approaches. In general, the own-price 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 1. The only elastic results were seen with the SUR-HEAR approach: the elasticity of sorghum acres with respect to corn and sorghum prices and the soybean elasticity with respect to the corn price. 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, e.g., ability to irrigate in dryland regions, topography, etc.

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, now it is the traditional approach that 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 average MSE with respect to the validation data set saw quick drops until reaching four hidden-layer neurons, the number of output neurons. Researcher burden could also be reduced through the inclusion of built-in marginal effect estimation for neural networks within statistical software.

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 a 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 that may be a little less familiar or even more difficult, such as artificial neural networks.

References

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).

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

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

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

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.

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

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.

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

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

Hendricks, N. P., Smith, A., & Sumner, D. A. (2014). Crop supply dynamics and the illusion of partial adjustment. *American Journal of Agricultural Economics*, *96*(5), 1469–1491.

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. (1990). Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural networks*, *3*(5), 551–560.

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).

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

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

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

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.

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.

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.

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

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`

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.

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.

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.

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, 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`

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.

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.
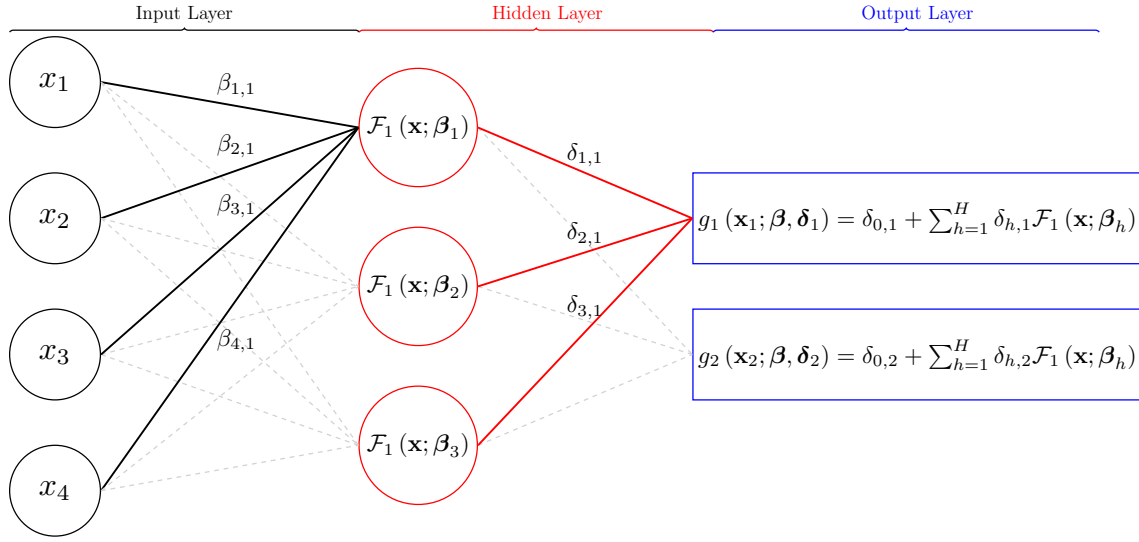
Figures
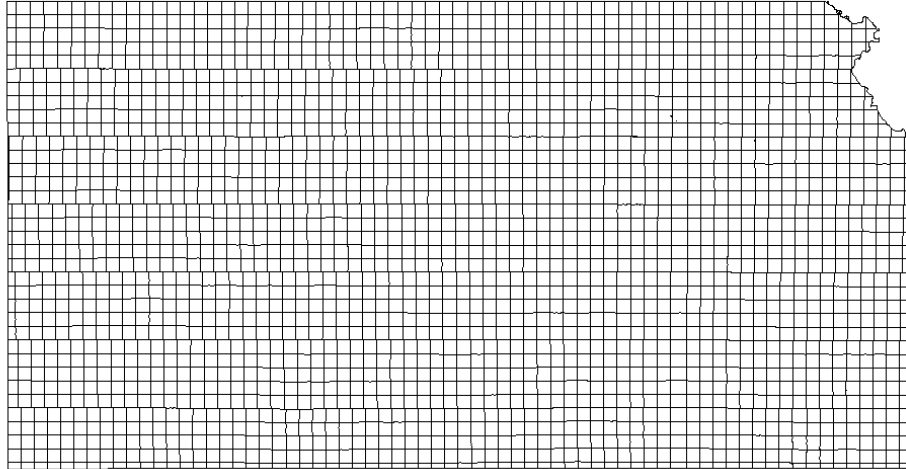

Figure 1: Log-acre-share ANN architecture

Figure 2: Kansas township divisions
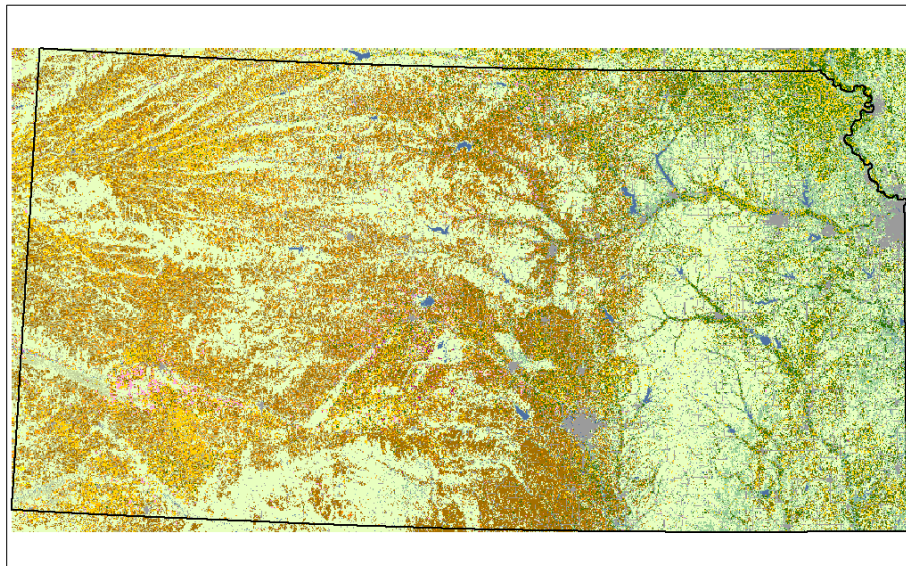
Figure 3: Cropland Data Layer example (2008)
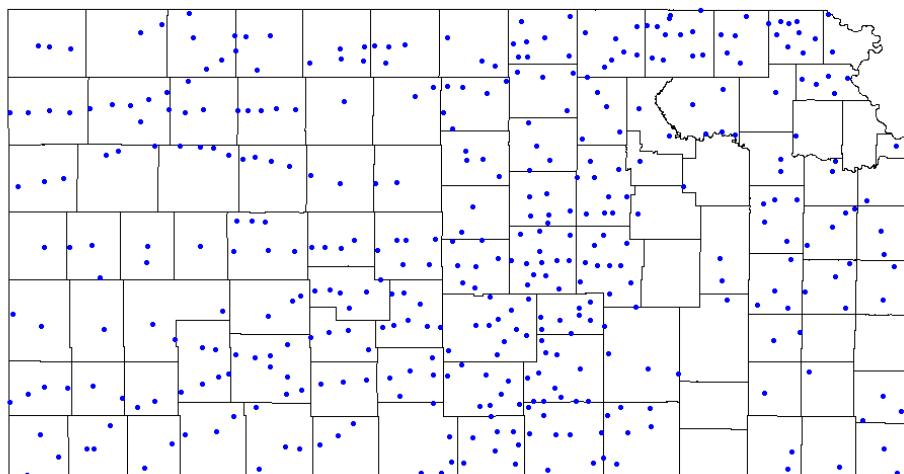
Figure 4: Elevator Locations
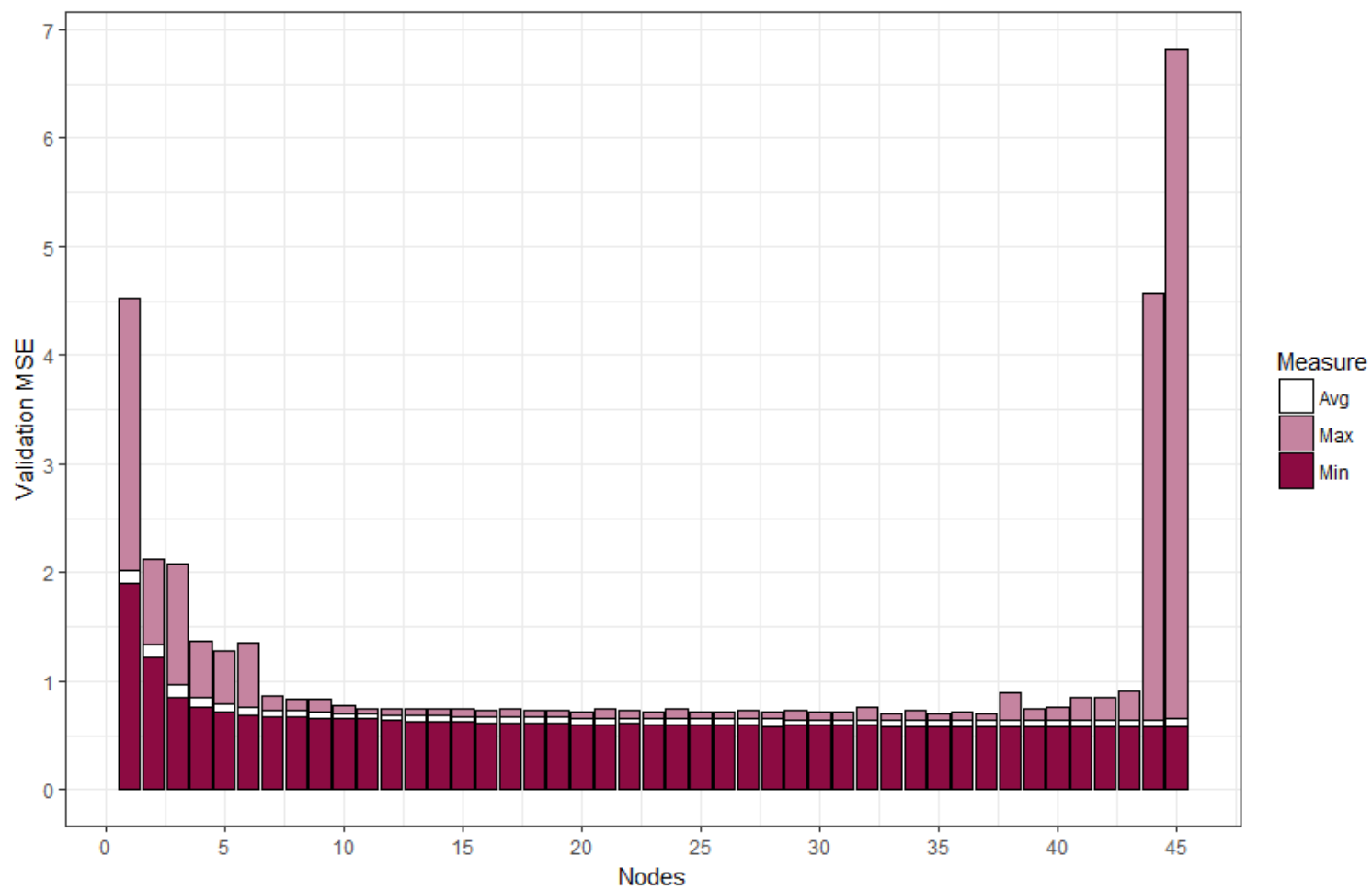
Figure 5: ANN performance across specifications
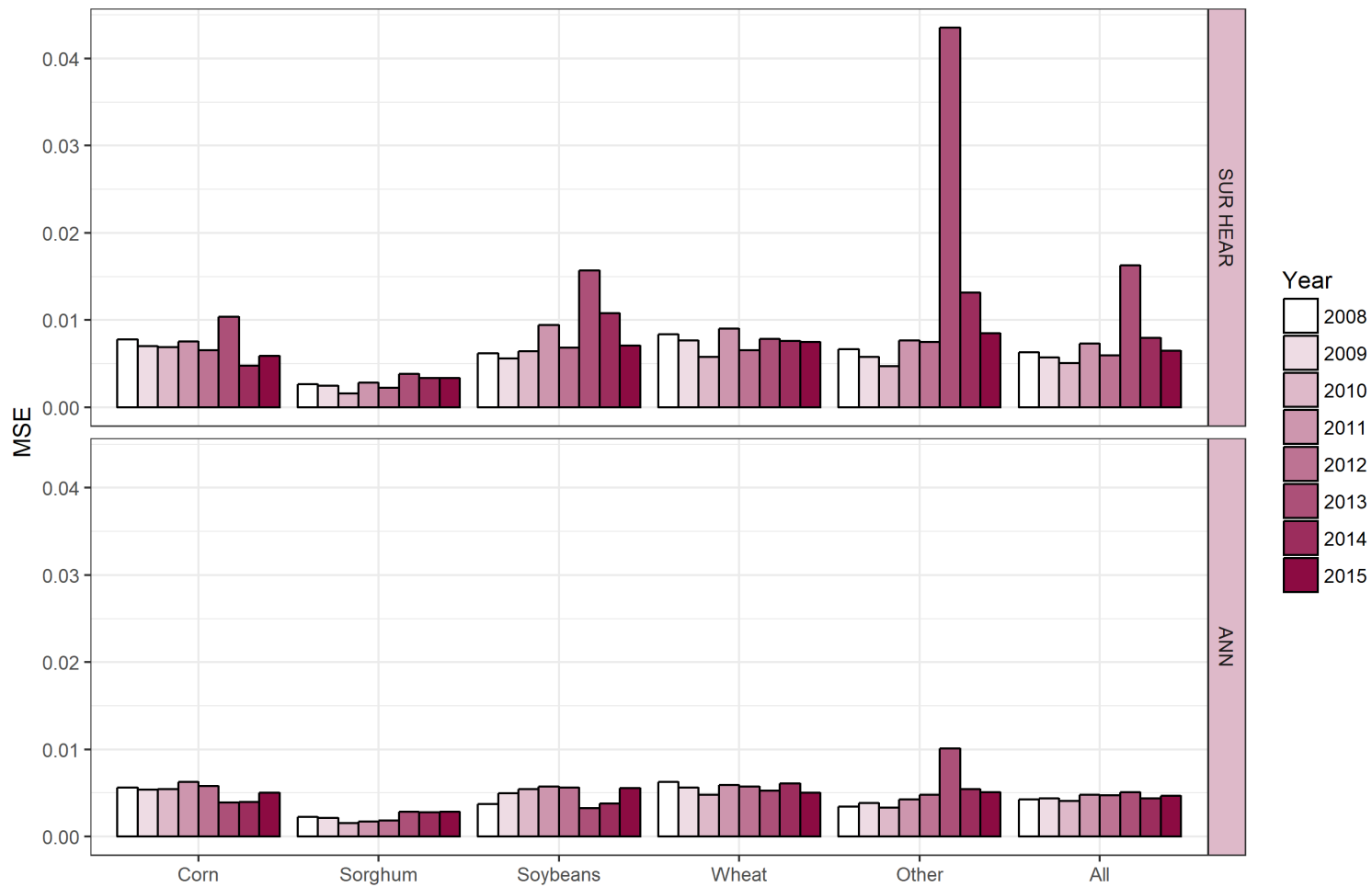
Figure 6: Mean square error
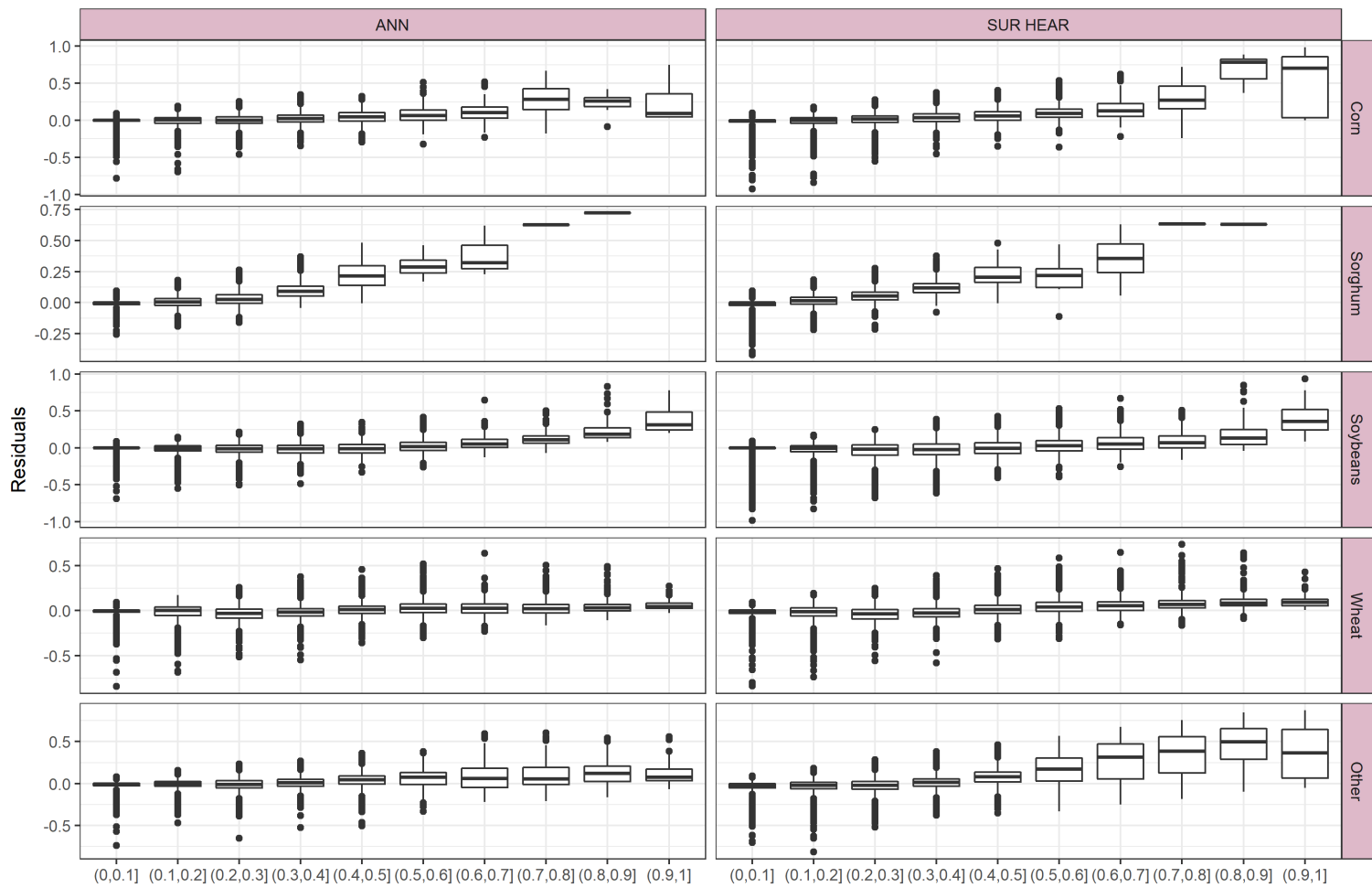
Figure 7: Residuals box-plots

Figure 8: Total acre predictions: Percent deviation from actual
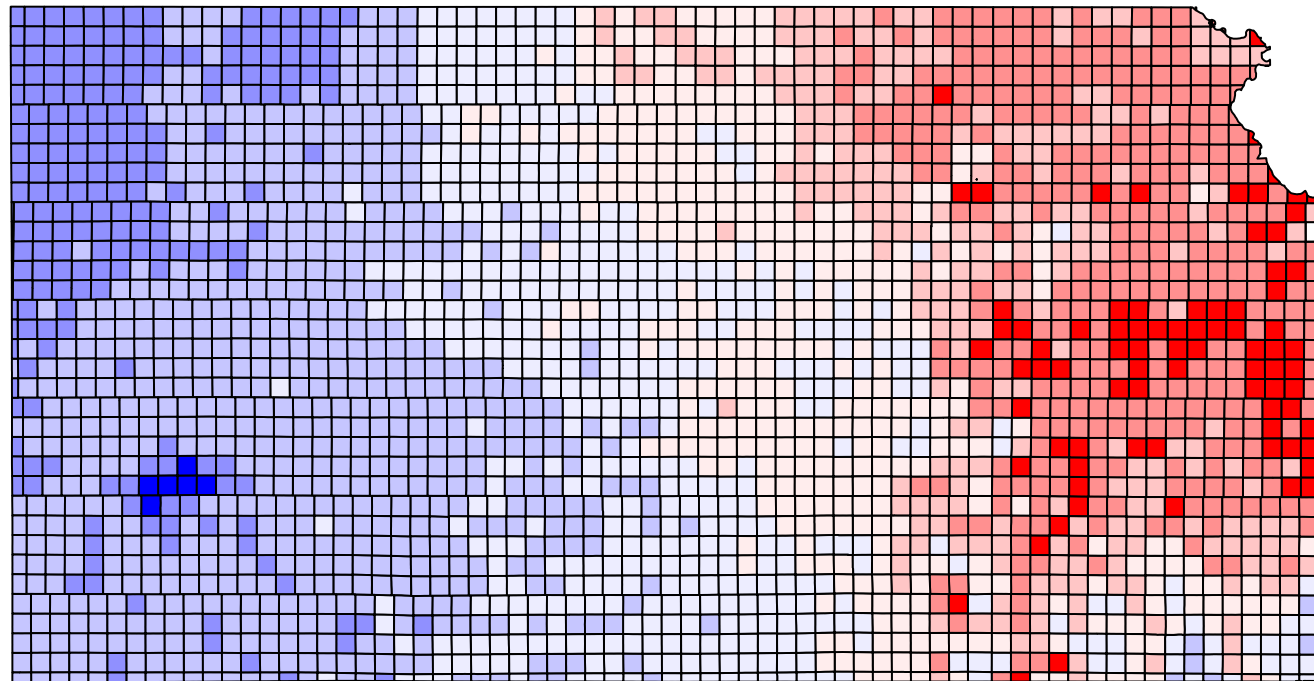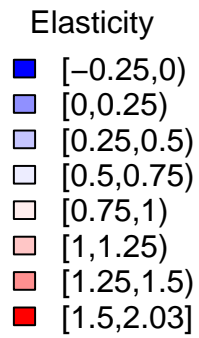
Figure 9: Corn own-price township-level elasticities, SUR-HEAR minus ANN

Figure 10: Sorghum own-price township-level elasticities, SUR-HEAR minus ANN
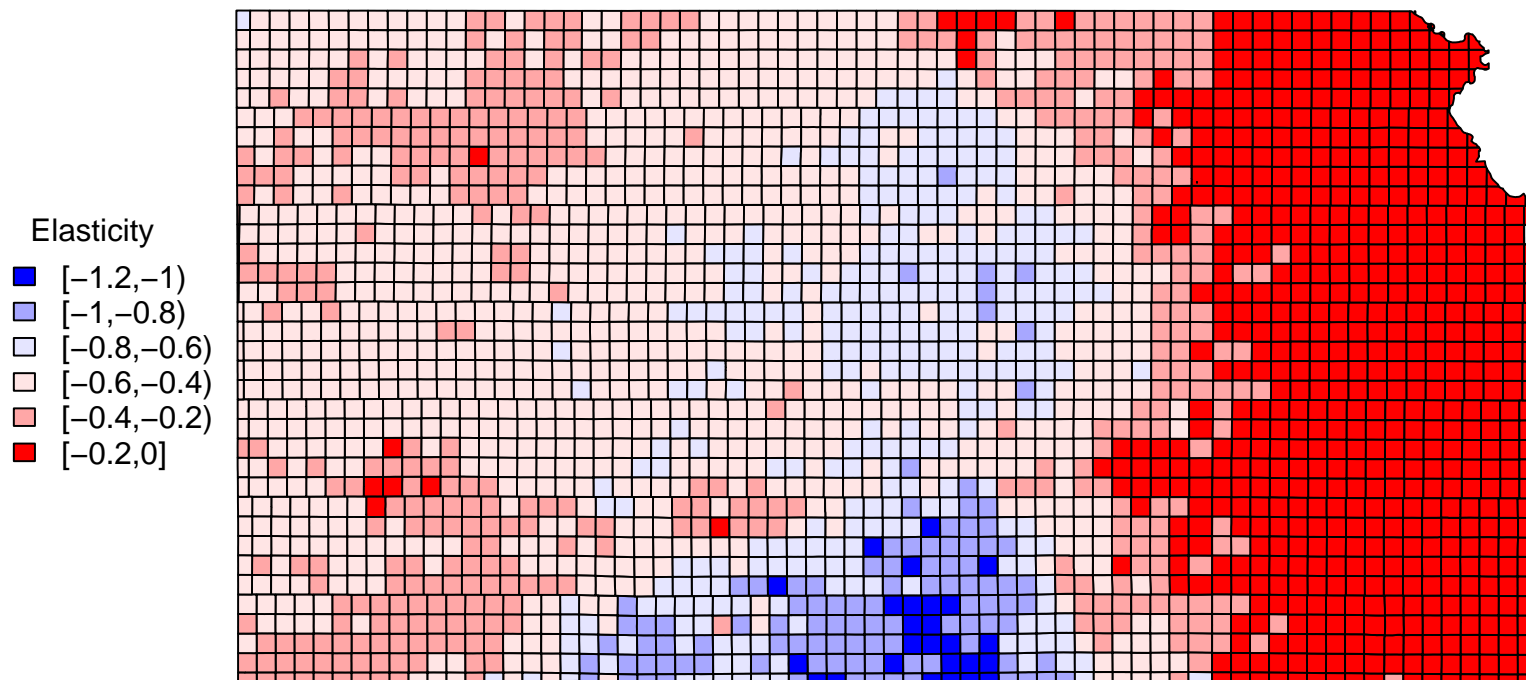
Figure 11: Soybeans own-price township-level elasticities, SUR-HEAR minus ANN
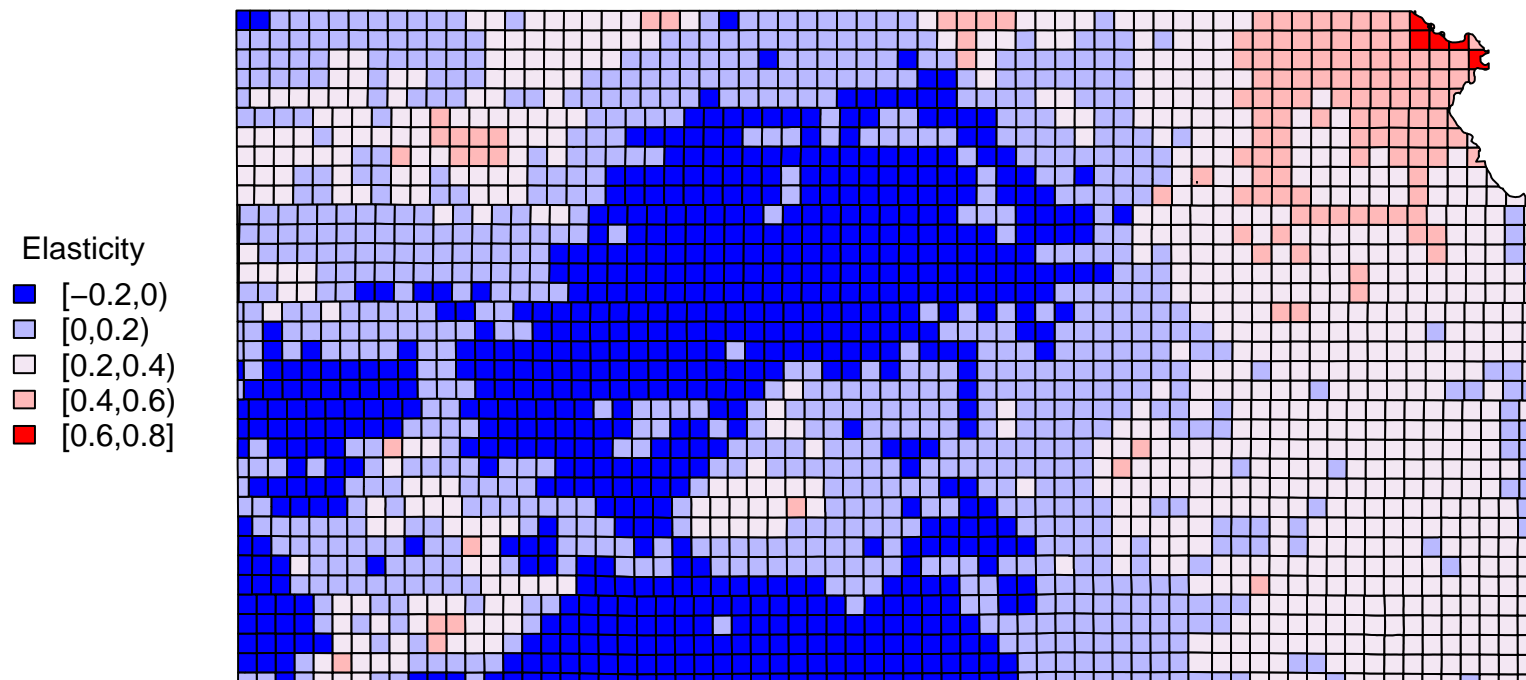
Figure 12: Wheat own-price township-level elasticities, SUR-HEAR minus ANN

Tables

Table 1: Acreage elasticity estimates from previous studies

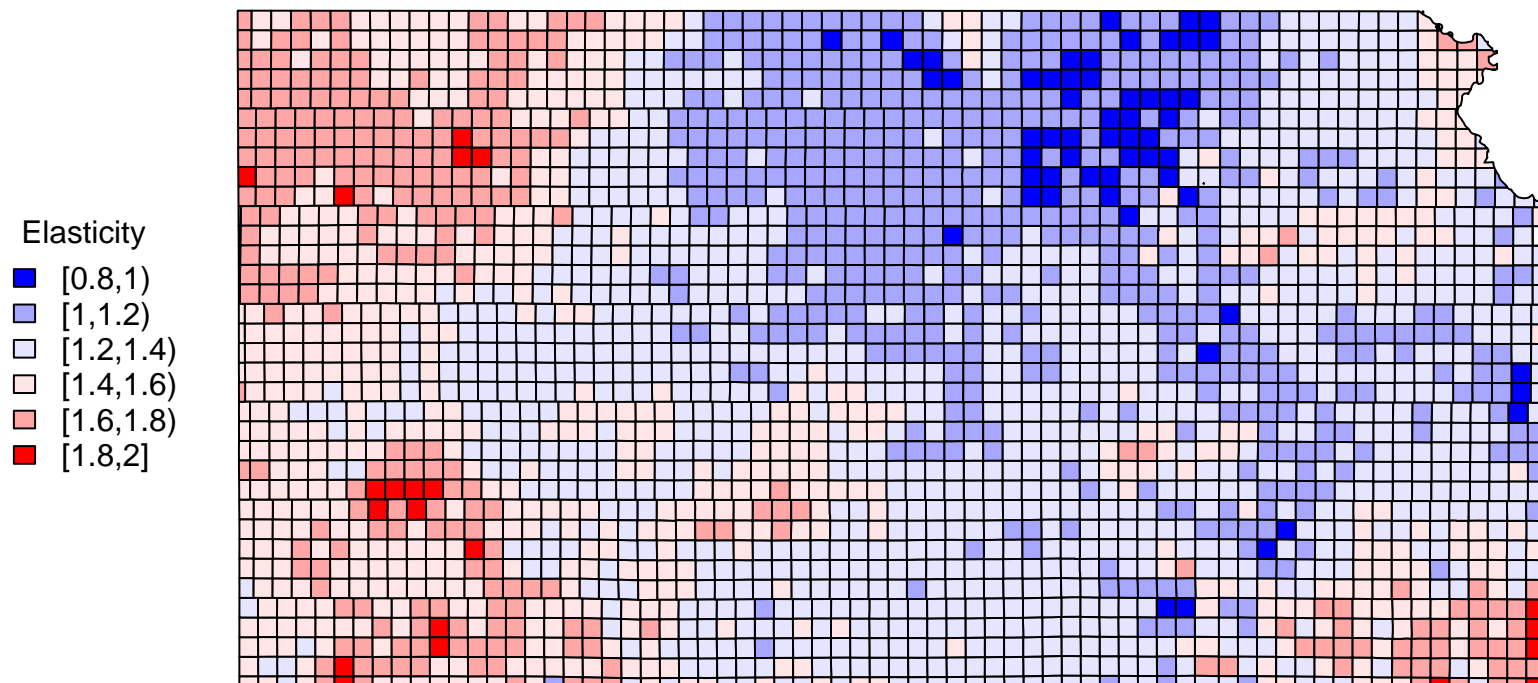| Study	Area	Time	Corn	Sorghum	Soybeans	Wheat
Adusumilli, Rister, Lacewell, et al. (2011)	Texas	1999-2009	0.06	-0.33	0.14	-0.12
Arnade and Kelch (2007)	Iowa	1960-1999	0.01		0.48	
Bailey and Womack (1985)	CO, KS, NE, NM, OK, TX, WY	1962-1981				0.25
Bridges and Tenkorang (2009)	NE, IL, IN, IA	1986-2007	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	
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 2: Summary data for land use shares

Year	Average Shares				
	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: Dependent and independent variables

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 4: SUR-HEAR test statistics

Equation	Test Statistic	Critical Value	p-Value	Decision
Autocorrelation, H_0: No autocorrelation				
Corn	0.467	2.706	0.10	Fail to reject H_0
Sorghum	0.372	2.706	0.10	Fail to reject H_0
Soybeans	0.596	2.706	0.10	Fail to reject H_0
Wheat	0.506	2.706	0.10	Fail to reject H_0
Heteroskedasticity, H_0: No heteroskedasticity				
Corn	16,295	2,503	0.01	Reject H_0
Sorghum	19,448	2,503	0.01	Reject H_0
Soybeans	10,840	2,503	0.01	Reject H_0
Wheat	20,804	2,503	0.01	Reject H_0
Contemporaneous correlation, H_0: No contemporaneous correlation				
All	23,424	16.81	0.01	Reject H_0

Table 5: Results for Ramsey RESET test

Equation	Variable	Coefficient p-Value
Corn	$\hat{CORN}_t^2$	0.948
Corn	$\hat{CORN}_t^3$	0.688
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	$\hat{WHT}_t^3$	0.003

Table 6: Mean-square error results, SUR-HEAR

Year	Corn	Sorghum	Soybeans	Wheat	Other	All
SUR-HEAR Model						
2008	0.008	0.003	0.007	0.007	0.007	0.006
2009	0.007	0.002	0.006	0.008	0.006	0.006
2010	0.007	0.002	0.007	0.006	0.005	0.005
2011	0.007	0.003	0.010	0.008	0.007	0.007
2012	0.007	0.002	0.007	0.006	0.007	0.006
2013	0.011	0.004	0.016	0.006	0.045	0.017
2014	0.005	0.003	0.011	0.008	0.013	0.008
2015	0.006	0.003	0.007	0.006	0.009	0.006
All	0.007	0.003	0.009	0.007	0.012	0.008
Average ANN MSE						
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 7: Deviations from actual acreages

	Corn			Sorghum			Soybeans			Wheat			Other		
Year	SUR-HEAR	BANN	WANN	SUR-HEAR	BANN	WANN	SUR-HEAR	BANN	WANN	SUR-HEAR	BANN	WANN	SUR-HEAR	BANN	WANN
2008	-0.10	-0.01	-0.03	-0.10	-0.01	-0.02	0.09	-0.02	0.02	0.03	0.00	0.01	0.02	0.02	0.01
2009	-0.01	-0.04	-0.02	0.04	-0.03	-0.06	-0.05	0.02	0.01	-0.03	0.01	0.02	0.09	0.02	0.00
2010	-0.03	0.01	-0.02	0.03	-0.04	-0.07	0.03	0.02	0.03	-0.03	-0.01	0.02	0.05	0.01	0.01
2011	-0.03	-0.04	-0.01	-0.12	-0.03	-0.08	0.11	0.01	0.03	0.01	0.01	0.01	-0.05	0.02	0.02
2012	-0.04	-0.08	-0.06	-0.16	0.00	-0.02	0.09	0.05	0.05	-0.03	0.02	0.01	0.07	0.01	0.00
2013	0.12	-0.05	-0.06	-0.19	-0.01	-0.02	0.49	-0.02	-0.06	0.05	0.00	0.02	-0.22	0.04	0.03
2014	-0.08	-0.04	-0.02	0.00	-0.04	-0.04	-0.31	0.03	0.03	0.02	0.01	-0.01	0.22	0.01	0.02
2015	-0.01	-0.03	-0.05	-0.16	-0.01	-0.01	0.01	0.00	0.00	0.02	0.02	0.02	0.04	0.00	0.02
MAD	0.05	0.04	0.03	0.10	0.02	0.04	0.15	0.02	0.03	0.03	0.01	0.01	0.09	0.02	0.01

Variable	SUR-HEAR				ANNs			
	Corn	Sorghum	Soybeans	Wheat	Corn	Sorghum	Soybeans	Wheat
$CORN_{t-1}$	0.37 *** (16.86)	0.02 ** (2.34)	−0.37 *** (−15.32)	−0.01 *** (−4.09)	0.01 *** (6.38)	0.00 (−1.21)	−0.01 *** (−3.33)	0.00 ** (−2.06)
SOR_{t-1}	0.10 *** (14.76)	−0.26 *** (−17.54)	−0.01 (−0.92)	−0.01 *** (−7.15)	0.01 *** (4.49)	−0.01 *** (−14.35)	0.00 (1.07)	0.00 * (1.74)
SOY_{t-1}	−0.36 *** (−18.33)	−0.11 *** (−8.48)	0.57 *** (21.59)	−0.04 *** (−9.01)	−0.01 *** (−3.49)	0.00 (−0.83)	0.02 *** (4.55)	0.00 (0.00)
WHT_{t-1}	−0.13 *** (−19.07)	−0.25 *** (−23.19)	−0.20 *** (−19.63)	0.15 *** (34.93)	−0.01 *** (−14.60)	−0.01 *** (−8.27)	−0.02 *** (−13.40)	0.02 *** (24.88)
$PCORN$	0.69 *** (6.59)	−1.03 *** (−6.88)	−1.56 *** (−14.48)	−0.03 (−0.77)	0.06 *** (2.86)	0.04 * (1.69)	−0.02 (−0.56)	−0.01 (−0.88)
$PSOR$	−0.75 *** (−8.33)	1.42 *** (11.74)	0.74 *** (8.65)	0.06 *** (3.93)	−0.03 * (−1.88)	0.04 (1.53)	0.01 (0.54)	0.00 (−0.34)
$PSOY$	−0.36 *** (−4.28)	−0.77 *** (−6.30)	−0.26 *** (−2.91)	0.40 *** (12.23)	0.00 (0.11)	−0.04 (−1.43)	0.03 (1.05)	0.01 (0.63)
$PWHEAT$	−0.69 *** (−16.95)	0.44 *** (8.37)	0.16 *** (3.69)	0.06 * (1.81)	−0.04 ** (−2.08)	0.04 (1.20)	−0.05 ** (−2.43)	0.02 (1.28)
$PDIESEL$	0.22 *** (2.70)	−0.63 *** (−5.45)	0.60 *** (7.30)	−0.29 *** (−7.65)	0.00 (−0.02)	0.00 (0.01)	−0.08 (−1.01)	−0.03 (−0.64)
$PLABOR$	0.15 ** (1.95)	0.69 *** (6.37)	−0.71 *** (−8.78)	0.05 (1.37)	−0.07 (−0.72)	0.08 (0.47)	0.05 (0.44)	−0.02 (−0.24)
AVG_{CSS}	0.10 *** (4.31)	−0.30 *** (−11.60)	0.77 *** (22.05)		0.01 (1.14)	−0.02 *** (−2.97)	0.01 * (1.69)	0.01 * (1.88)
AVG_W				−0.19 *** (−17.36)	−0.02 *** (−3.83)	0.01 (0.90)	0.01 (1.00)	0.00 (−1.15)
$PREC_{CSS}$	0.01 (0.80)	−0.01 (−0.50)	0.19 *** (10.30)		0.01 *** (3.39)	0.01 *** (2.68)	0.00 (−1.06)	0.00 ** (−2.16)
$PREC_W$				−0.01 *** (−3.28)	0.00 (0.38)	0.00 (−0.49)	0.00 (0.72)	0.00 (−0.21)
$TMAX_{CSS}$	1.31 *** (12.41)	1.21 *** (7.24)	0.21 * (1.89)		0.04 (0.99)	−0.05 (−0.72)	−0.04 (−0.61)	0.03 (0.86)
$TMAX_W$				−0.24 *** (−5.45)	−0.06 (−1.45)	0.01 (0.08)	−0.21 *** (−4.23)	0.03 (0.98)
$TMIN_{CSS}$	−0.78 *** (−15.50)	−0.68 *** (−9.99)	−0.01 (−0.10)		−0.03 (−1.64)	−0.04 (−1.26)	0.06 * (1.77)	0.00 (0.01)
$TMIN_W$				0.15 *** (11.04)	0.01 (0.32)	0.02 (0.58)	0.00 (−0.13)	0.01 (0.99)

Values in parentheses denote z-statistics

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