IOWA STATE UNIVERSITY
Digital Repository

2015

Three studies on applying Positive Mathematical Programming and Bayesian Analysis to model US crop supply

Michael Hudak
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/etd

Part of the Agricultural and Resource Economics Commons, and the Agricultural Economics Commons

Three studies on applying Positive Mathematical Programming and Bayesian Analysis to model US crop supply

by

Michael Hudak

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Economics

Program of Study Committee:
Bruce A. Babcock, Major Professor
Chad E. Hart
Peter F. Orazem
Sebastien Pouliot
Stephen B. Vardeman

Iowa State University

Ames, Iowa

2015

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank my committee chair, Dr. Babcock, and my committee members, Dr. Hart, Dr. Orazem, Dr. Pouliot, and Dr. Vardeman, for their guidance and support throughout the course of this research.

In addition, I would also like to thank my friends, colleagues, the department faculty and staff for making my time at Iowa State University a wonderful experience. I want to also give thanks to my family for their encouragement in my pursuit of a doctorate.

GENERAL ABSTRACT

The purpose of this dissertation is to find a practical way of obtaining a reasonable crop supply model for the US using a limited dataset. This model can then be used for forecasting and impact modeling. The method that is central to this model is Positive Mathematical Programming (PMP) that allows for the calibration of a nonlinear programming model to mimic the observations. This method is improved by implementing Bayesian Analysis to allow for the model to consider a distribution for the supply elasticity.

Using this method a national model was formed using only five years of data. While there were difficulties in forming a posterior density through manipulation of parameters, the Metropolis Hastings Algorithm ultimately allowed for the density to be simulated. Once the posterior data is simulated, a reasonable forecast could be made using this model.

This model was then improved by disaggregating the national model into a regional model. This was done through an additional variable (which is the percentage of national price responsiveness for a crop in a region) to consider in the prior density. Ultimately, regional results and elasticities are formed and the overall forecasting was improved.

Once the national and regional models have been formed, the models were tested under a variety of impact models. The response to the change in price for crops as well as yield changes in a region were done and reasonable results were found. Overall, a crop supply model was formed that produced reasonable elasticities and forecasted accurate results, thanks in part to a Bayesian approach which view parameters as distributions in the model.

CHAPTER I: ESTIMATING US CROP SUPPLY MODEL USING PMP AND BAYESIAN ANALYSIS

1.1 Abstract

This chapter develops an innovative and practical way to model the supply of agricultural crops. This is done by extending the technique developed by Howitt (1995), Positive Mathematical Programming (PMP), using Bayesian estimation. A key problem in the use of the PMP model is the relative difficulty of finding calibrating parameters such that the first and second order conditions are satisfied. There is the added difficulty that many of the conditions needed to be satisfied are not exactly known. Thus the use of Bayesian analysis is a useful tool to help determine these parameters. By employing a Markov chain Monte Carlo (MCMC) algorithm, specifically a Metropolis-Hasting algorithm, a posterior distribution for the calibrating parameters can be found such that the resulting supply model will not only create an optimum that reproduces observed acreages, but will also produce reasonable elasticities in accordance with prior information. The value of this style of estimation for a crop supply model lies in the limited amount of data needed to estimate the model.

1.2 Introduction

The purpose of this work is to provide a new way to estimate an agricultural supply model. There are many reasons that crop supply models are useful. First, they can be used to test hypotheses regarding the impacts of policy changes on the supply of various crops. Second, these models can be utilized to forecast the supply of various crops. There are two methodologies that are useful in achieving a quantitative model for the supply of various crops. The method used in

much of the literature involving the estimation of supply is the dual equation system which can specify the model with well established econometric techniques. The core idea of dual equations is to utilize the profit function of the various crops together with the various input demand equations that normally create a variety of resource restrictions, even if the overall model is unconstrained. The disadvantage of this approach is that it often oversimplifies the model and normally requires a lot of input data in order to be estimated. An example of this is given by Chambers and Just (1989) who employ a dual supply model with explicit allocations for fixed factors. The second methodology is to use a programming model which can model complex problems. The programming methodology has provided practical models for crop supply, however, it usually lacks statistical properties. One such-programming model which has found a great deal of success was developed by Howitt (1995), who devised Positive Mathematical Programing (PMP).

The basic idea of PMP is to form a non-linear programming problem for land allocation in a crop supply model by adding a quadratic portion to a known linear programming problem for crop supply. The key is that the added quadratic portion will calibrate the model such that it cannot only reproduce an optimal solution that will be consistent with observations of land allocations, but will also calibrate the model so that it is consistent with economic theory. Thus, this model is consistent with the fact that the cost curve is convex. Additionally, the model is calibrated to an assumed correct belief of what the shadow price of land should be, and it properly reflects the assumed correct price elasticity for the various crops. It is essential that the model developed is consistent with other observations in the literature with respect to price elasticities, as the relevancy of this modeling is how well it can predict changes due to shifts in the supply curve. The benefits to PMP are that the policy analyst can easily form a crop supply

model that is not only calibrated to the observed data, but also does not need a lot of data for its

calibration. The present research extends the PMP method by applying Bayesian estimation in

order to not only estimate a crop supply model, but to also derive a crop supply model that is

testable, can be forecast, and can simulate a wide range of policy changes.

The work builds upon previous work done by Jansson and Heckelei's (2011) European

Union (EU) model, but with two key differences. First, the present application is made using US

data, which adds some additional challenges in modeling. Second, the Metropolis Hasting

algorithm is used to estimate a model that provides the creation of credible intervals and other

forms of Bayesian inference not present in previous research in this area. With these estimations,

forecasting supply as well as the effects of policy becomes feasible.

1.3 Literature Review of PMP

The evolution of estimating agricultural supply models in a practical way is best seen

though Heckelei and Britz's (2005) examination of PMP. The standard model considered is the

simple linear problem:

$$\max_{x} Z = p'x - c'x$$

$$\text{s.t. } Rx \leq v \text{ and } x \geq 0$$

(1)

Here p is a vector of revenues for various crops, x is a vector of acreages for the various

crops, c is linear cost vector for the various crops, R is a matrix of coefficients in the resource

constraints, and v is a vector of the various available resource quantities. The drawback of this

model is that it normally does not predict the observed value for x and it typically predicts over

specialization in crops, because the model does not have decreasing marginal profits for any of

the crops. This omission is a critical flaw of the model. The overall solution to this standard

model would be specialization into a few crops, since usually the number of constraints are less than the observed activities. PMP was in part developed to create a model whose optimal solution could reproduce what is observed and would fit the standard properties of a profit function. The contribution of PMP by Howitt (1995) was to consider an additional non-linear portion to the objective function allowing for the model to have decreasing marginal profits in all the crops. The non-linear model can then be represented as:

$$\max_{x} Z = p'x - c'x - h'x - \frac{1}{2}x'Qx$$

(2)

$$\text{s.t. } Rx \leq v \text{ and } x \geq 0$$

The main concern in this non-linear function is that the parameters in Q and h are unknown. The only real restriction on the parameters is that Q has to be a positive semi-definite matrix in order to be consistent with standard economic theory of profit maximization. There are a number of ad-hoc methods that have been used to try to model this non-linear portion. In most cases these methods involve adding some element, such as risk, into the modeling. For example, Holt (1999) formulated the problem as a portfolio choice problem. However, a more practical solution usually used with the PMP method is to assume these parameters are unknown and use observed data to calibrate them.

By simply using one year of data involving observed acreage, prices, and crops, the model can be calibrated. Thus the optimal solution for x is the acreage observed which is ensured provided that the first order conditions of the model at the observed acreage are met. For this condition to occur, various artifices are employed. If the resource constraint only pertains to the land constraint, the simplest solution is to assume the shadow price of land is equal to the average profit of an additional unit of land across all the activities of the model. With this

assumption the parameters of h and Q could be defined in such a way that the optimal solution replicates what is observed. The only problem with such an approach is that the first order conditions create an ill-posed problem: that is, multiple combinations of h and Q will solve the equation. Howitt's initial modeling used a simple specification for h and Q which could be used to calibrate the model to be optimal with respect to the observed acreages. However, the model could not accurately predict the optimal outcome if prices shift. What is needed is applying an additional restriction to the model using some known piece of information.

Heckelei and Britz (2005) show from the first order condition, that the optimal acreage x^* can be found. In addition, by taking the derivative of x^* with respect to price, $\frac{\partial x^*}{\partial p}$ is found and from this the parameters Q can be equated to the various crop elasticities which are (in some cases) assumed known. This additional information allows for the parameters of h and Q to have a unique set of outcomes for which the crop supply model will not only reproduce the observed acreage, but also reproduce an observed crop elasticity.

There are, however, two major problems associated with using exogenous elasticities. The first is that it is rare to exactly know what a crop elasticity is in a given year. The other complication lies in the fact that solving for what Q should be in order to reproduce exactly the exogenous supply elasticity is difficult and in some cases infeasible, as shown by Merel and Bucaram (2010). Merel and Bucaram (2010) also raise the question of whether or not the calibration to observed acreage data is unique, which would ultimately need to be addressed in a PMP model. Because of the difficulties of exact calibration, a popular idea has emerged to use multiple observations to help estimate the parameters of the model by using a Maximum Entropy (ME) approach.

The first use of ME techniques to solve this problem was done by Paris and Howitt (1998). However, their use of these techniques was rather limited in that they simply used one observation in their study. Heckelei and Wolff (2003) examined the performance of a ME estimator though Monte Carlo simulations with small samples and found a general improvement in using a maximum entropy approach. The key idea of ME is that instead of making a specific parametric assumption, one assumes a uniform distribution for the parameters. Moreover, the parameters must satisfy the given restrictions implied by the first-order conditions while being as close to the uniform distribution as possible. By using ME, time series data could be included. There is a great deal of flexibility with the ME approach and this method opens up the possibility to calibrate the model uniquely without the need for knowledge of the elasticities.

There have been numerous articles that employ the PMP approach together with various ways of estimating the unknown parameters. On such approach was Kanellopous et al. (2010) with the Farm System SIMulator. They assessed the forecasting performance of a PMP model. They needed to do several modifications with respect to how the shadow price is estimated, apart from the standard method given by Howitt, in order for the approach to work with time series data.

There has also been work done by Buysse et al. (2007) in the estimation of the cost function of Belgian sugar beet through the use of a Generalized Maximum Entropy method on a farm programming model and an additional problem unique to the model of sugar beet quota rents. Other work has been done by Arfini et al. (2008) using the PMP estimation in a cross sectional problem in the Emilia-Romangna region of Italy. The calibration was done not only for the production of crops but also for the demand function of the crops which allowed the model to have endogenous prices. Another use of PMP modeling has been done by Gocht (2009), using

PMP to estimate input allocations with the claim that this style of estimation can lead to more information on the input coefficients than previous estimation approaches.

There is, however, an alternate way to define the parameters of the model while still using a limited amount of information. Instead of the ME approach, a Bayesian approach could be used, as explained by Heckelei (2008). A Bayesian approach can recreate a ME approach, but is more flexible in incorporating prior information in the model, while also being simpler to compute.

This paper uses a model similar to that used by Jansson and Heckelei (2011) who utilized a Bayesian estimation method on time series data to estimate the supply model for the EU. The model used by Jansson and Heckelei is an errors in variables model. The benefits to using a measurement error model is explained by Carroll et al. (1995). They demonstrate a general improvement to the estimations done through Gibbs sampling when taking into account measurement error. In addition, this type of modeling is useful for filling in missing data by treating unobserved X's as an unobserved random parameter with a distribution based on some prior distribution. The ability to deal with incomplete data is a major reason to consider using an errors in variable model. While the general model this thesis poses could allow for errors in variable, we assume for simplicity some variables are precisely measured. Thus in the empirical portion, there will be no measurement error in measuring the profit per acres of crops. This assumption is implemented primarily to limit the number of parameters that are needed to be determined. Jansson and Heckelei's (2011) basic modeling and approach could be adapted to the US, but noticeable changes are necessitated when modeling a single country's agricultural supply model with information existing in subregions. There is also the problem that while Jansson and Heckelei (2011) could estimate a model using Bayesian estimation, they could not overcome the

problem of determining the significance of the estimation since they used a form of optimization rather than an estimation of the parameters posterior distributions. Because of this, the method used in my paper employed several alterations to allow for a posterior distribution to be sampled using an MCMC algorithm, a Metropolis-Hasting Algorithm. By using an alternative estimation method, we are not only able to estimate the unknown parameters, but also will be able to give credible intervals and thus some significance to the estimation. First a review of the generalized model posed by Jansson and Heckelei is needed.

1.4 US Crop Supply Model

The core of the basic model used by Jansson and Heckelei (2011) is Howitt's (1995) PMP model using time series data:

$$\max_{x_t} x_t[p_t Y_t - A_t w_t] - q_t x_t' \left[h + \frac{1}{2} l_t Q x_t\right]$$

(3)

$$\text{s.t. } R_t x_t = v_t$$

Where x_t is a vector of acreages for J crops, Y_t is a J×J matrix of yields, p_t is a J vector of prices, A_t is a J×N vector of input coefficients for N inputs, w_t is an N vector of input prices. Q is a J×J diagonal matrix of own crop effects to be estimated. The cross crops effects in Q were not considered for this model for several reasons. First, accurate information relating to direct interaction between crops in terms of marginal cost is difficult to find. The only information known for the supposed synergy between crops for corn, soy and wheat, is that if cross effects do exist, it would be a magnitude smaller than the crops own impact on marginal cost. With that fact we can assume the form for Q should be approximately diagonal. This assumption greatly facilitates the computation and determination of the posterior distribution, as explained in a later

section. The only constraint considered will be the land constraint. Thus R_t is a J vector of ones and v_t is total land available. This model is a very general PMP model. $[P_t Y_t - A_t w_t]$ represents the observed marginal profits per acreage of crop j. $q_t x_t' \left[h + \frac{1}{2} l_t Q x_t \right]$ is just the PMP portion added in to be estimated. The use of q_t is a general price index employed to deflate opportunity costs, and the scalar l_t is used to prevent the possibility of land migration from influencing land rent. The parameter h is still an unknown linear portion in the problem and Q is the unknown quadratic portion. The first order and second-order conditions are then:

$$p_t Y_t - A_t w_t - q_t h - q_t l_t Q x_t - R_t \lambda_t = 0 \tag{4}$$

$$R_t x_t = v_t \tag{5}$$

$$Q = U'U \tag{6}$$

Where $Q = U'U$ ensures that the quadratic portion of the behavioral matrix is positive semidefinite. Thus the additional constraints to be considered when using time series data are as follows.

Let $E_t = q_t l_t Q$, and thus its inverse being E_t^{-1}. Solving for (4) it can be found that

$$x_t^* = E_t^{-1}[p_t Y_t - A_t w_t - q_t h - R_t \lambda_t^*] \tag{7}$$

Substituting into (5) and solving for λ_t^* gives:

$$\lambda_t^* = [R_t E_t^{-1} R_t']^{-1}[R_t E_t^{-1}(P_t Y_t - A_t w_t - q_t h) - v_t] \tag{8}$$

Substituting (7) into (4) then gives

$$x_t^* = E_t^{-1}[p_t Y_t - A_t w_t - q_t h] - E_t^{-1} R_t([R_t E_t^{-1} R_t']^{-1}[R_t E_t^{-1}(p_t Y_t - A_t w_t - q_t h) - v_t]) \tag{9}$$

Taking $\frac{\partial x_t^*}{\partial p_t}$ gives

$$\frac{\partial x_t^*}{\partial p_t} = E_t^{-1}Y_t - E_t^{-1}R_t[R_tE_t^{-1}R_t']^{-1}R_tE_t^{-1}Y_t \tag{10}$$

From here η_t could be found

$$\eta_t = X_t^{-1}(E_t^{-1}Y_t - E_t^{-1}R_t[R_tE_t^{-1}R_t']^{-1}R_tE_t^{-1}Y_t)P_t \tag{11}$$

X_t, Y_t, and P_t all represent diagonal matrices of x_t, y_t, and p_t respectively. This elasticity is

consistent with the previous literature when considering the fact that:

$$\frac{\partial \lambda_t^*}{d\partial} = E_t^{-1}R_t[R_tE_t^{-1}R_t']^{-1}R_tE_t^{-1}Y_t. \tag{12}$$

In some cases it may be simpler to just consider $\frac{\partial \lambda_t^*}{\partial p_t} = 0$ and solve for the behavioral estimates.

While this model is similar to that posed by Jansson and Heckelei (2011), the Bayesian

estimation of this model is what makes this model unique. Jansson and Heckelei's model

attempted a Bayesian approach to the problem, however, caveats to their model were due to the

priors not truly being a pre-data view of the problem. As a result, they were only able to produce

the highest point for the posterior density, which gives no information as to the significance of

this point estimator. In fact, there is no proof that this highest point is unique, and there is no way

of knowing whether or not the distribution has a single peak. It is with these issues in mind that

one major goal of this research is to extend the Bayesian setting to US data in such a way that

credible intervals for the estimated parameters will be obtained. A major reason why Jansson and

Heckelei did not apply an MCMC algorithm is due to the size of their problem impacting the

speed of the algorithm. The model this paper will use has the advantage of focusing on US crop

supply rather than all of the EU's crop supply model. This model, however, will be scaled down

slightly to include fewer range of parameters and excludes measurement errors in some

variables. Utilizing a MCMC algorithm is a major benefit in a Bayesian setting. Since the only

information obtained by Jansson and Heckelei's approach is this highest point, there is no

calculation of the overall significance, or the posterior distribution. However, with the use of a

Metropolis-Hasting algorithm, not only is a posterior distribution for the parameters obtained,

but also a sample of the posterior is found making policy analysis feasible. This type of modeling

also has a significant benefit compared to linearly approximating the supply curve. With linear

approximation, only one optimum is normally considered for the problem, and a single peak is

assumed for its distribution, which does not need to be the case. In addition, without including

any prior information on the elasticity, trying to linearly approximate the supply curve using a

limited amount of this data would necessitate a huge range as to what the linear approximation

could be. By imposing a prior, the range that the supply curve could lie is reduced.

1.5 Bayesian estimation and Metropolis-Hasting Algorithm

The first goal in the estimation is to properly set up a model with reasonable priors and

likelihood densities. To simplify the model, let θ be the set of parameters that define the

unobservable quadratic portion of the problem, or $\theta = (h, U)$ where $D = U'U$. Another

simplification is to condense the observations from $(y_t p_t)$ and $(A_t w_t)$ into simply $revenue_t$ and

$cost_t$ which are J vectors of the average revenue or cost per acre. As for the data, we are able to

observe x, revenue, and cost, and from that we also infer the shadow price λ which is an average

of the observed gross margins of all the crops. Thus the observed data is $z =$

$(x^{obs}, revenue^{obs}, cost^{obs}, \lambda^{obs})$. Let us also define $x^* = x(\theta)$, $\lambda^* = \lambda(\theta)$, and $\eta^* = \eta(\theta)$, For

times series these would be:

$$x_t^* = E_t^{-1}[revenue_t - cost_t - q_t h]$$

$$- E_t^{-1} R_t ([R_t E_t^{-1} R_t']^{-1}[R_t E_t^{-1}(revenue_t - cost_t - q_t h) - v_t]) \tag{13}$$

$$\lambda_t^* = [R_t E_t^{-1} R_t']^{-1}[R_t E_t^{-1}(revenue_t - cost_t - q_t h) - v_t] \tag{14}$$

and

$$\eta_t^* = diag(x_t^*)^{-1}(E_t^{-1} - E_t^{-1} R_t [R_t E_t^{-1} R_t']^{-1} R_t E_t^{-1}) diag(revenue_t) \tag{15}$$

The goal is to obtain a posterior density, $g(\theta|z)$, which is based on observations

$z = (x^{obs}, revenue^{obs}, cost^{obs}, \lambda^{obs})$. The basic Bayes rule should now be:

$$g(\theta|z) \propto f(z|\theta)g(\theta) \tag{16}$$

Here the biggest question is how to form a prior for $g(\theta)$. Available information to form

the prior is the estimates in the literature of supply elasticities which can basically be seen as a

function of θ or $g_\eta(\eta(\theta))$. The basic information for the distribution $g_\eta(\eta(\theta))$ is that the mode

of the distribution should be around 0.25 for corn and soy and 0.15 for wheat. Hendricks, Smith,

and Sumner (2014) gives a general range for corn and soy elasticities and .25 was decided to be a

good place to try and center the distribution for the elasticity of corn and soy. Wheat in general is

seen as more inelastic in terms of supply so 0.15 was picked for it. Overall these distributions

work as long as the range of the distribution is from 0 to 1. From this information and given

average data for x^{obs} and expected price for the crop, this prior belief can be transformed into a

prior belief on $\frac{dx^*}{dp}$ so then the prior belief we have is $g_{slope}\left(\frac{dx^*}{dp}(U)\right)$. In trying to form a

distribution for $g_{slope}\left(\frac{dx^*}{dp}(U)\right)$ we use the following rule. The mode of the distribution is

found such that it is the mode of the average elasticity over the five years being observed. Thus

the distributions for $\frac{dx^*}{dp}$ have modes at 4.5 million acres for corn, 1.5 million acres for soy, and 1

million acres for wheat. Thus what is needed is a distribution for g(θ) such that draws from that

distribution would simulate $g_{slope}\left(\frac{dx^*}{dp}(U)\right)$. One key assumption that makes this problem

easier is that Q, and by extension U, are diagonal matrices. This almost guarantees a solvable

functional form relating $\frac{dx^*}{dp}$ and U. This functional form is developed and described in Chapter 2.

Measuring at the mean of the data, it is found that $g(U_{11}){\sim}N(3077,(700)^2)$,

$g(U_{22}){\sim}N(3808,(700)^2)$, and $g(U_{33}){\sim}N(2914,(700)^2)$ works to simulate the slope. A

variance of $(700)^2$ was chosen since it gave a reasonable range of possible elasticities without

yielding extreme ranges. A different range of U_{ii} can be experimented with when employing an

alternative prior, as is done in a later section for a robustness check. A prior still needs to be

specified for $g(h)$. What is needed is a distribution to be conditional on U, or $g(h|U)$. This was

done because for a given U, there is a limited range on h, such that the over all results for $x(\theta)$ is

in a feasible range. The conditional function is $h_i{\sim}N(h_i(U_{ii}),25)$. The function form for $h_i(U_{ii})$

can be found by altering the first order condition in terms of l_i giving:

$$h_i = \frac{\frac{mean(x^{obs})mean(q)}{U_{ii}} - mean(revenue^{obs}) + mean(cost^{obs}) + mean(\lambda^{obs})}{-mean(q)} \quad (17)$$

Thus the mean of the described distribution is centered on what would be the most

feasible point given U_{ii}, and a variance of 25 which is wide enough to consider alternative values

for h_i. It should also be noted that U_{ii} and h are calibrating unknown parameters, so exactly

knowing what these are measuring is not possible. The overall units of measure for h_i would be

in dollars per acre, while Q_i would be in dollars per acre squared, and thus elements of U_{ii} would

be the square root inverse of that.

Next the likelihood function is $f(z|\theta) = f(x^{obs}, revenue^{obs}, cost^{obs}, \lambda^{obs}|\theta)$. In order to construct this likelihood function, there needs to be a strong assumption that the measurement errors between various observations are independent, so that the likelihood function could be redefined as:

$$f(z|\theta) = f_x(x^{obs}|\theta)f_\lambda(\lambda^{obs}|\theta)f_{rev}(revenue^{obs}|\theta)f_{cost}(cost^{obs}|\theta) \qquad (18)$$

Since there is an assumption that there is no measurement errors in revenue and cost, it can be assumed that $f_{rev}(revenue^{obs}|\theta) = f_{cost}(cost^{obs}|\theta)=1$.

Thus the only part of the distribution that has relevance to the likelihood function is $f_x(x^{obs}|\theta)f_\lambda(\lambda^{obs}|\theta)$. $f_x(x^{obs}|\theta)$ is a joint truncated normal distribution such that. $f_x(x^{obs}|\theta) = \prod f_{x_j}(x_j^{obs}|\theta)$, where j represents the three crops. Each of the distributions $f_{x_j}(x_j^{obs}|\theta)$ are truncated normal with mean at $x_j(\theta)$, truncated at 0 and with standard deviation being $\sigma^2 = (.2/3 * \mu_{x_j})^2$. This formulation is used because small changes in θ tend to have large impacts on $x(\theta)$. Thus in order for a MCMC algorithm to work efficiently, a higher standard deviation on the measurement is assumed. $f_\lambda(\lambda^{obs}|\theta)$ is assumed to be normally distributed, truncated at 0 with a standard deviation equal to 25% of λ, that is:

$$f_\lambda(\lambda^{obs}|\theta) = \frac{\varphi\left(\dfrac{\lambda^{obs} - \lambda(\theta)}{.25 * \lambda(\theta)}\right)}{.25 * \lambda(\theta)\left(1 - \Phi\left(\dfrac{-\lambda(\theta)}{.25 * \lambda(\theta)}\right)\right)}$$

All that is needed is to find $g(\theta|z)$ based on these distributions. The method used here to find $g(\theta|z)$ is an MCMC algorithm, namely the Metropolis-Hasting algorithm. There are several instances in the literature of using an MCMC algorithm to obtain posterior distributions. An MCMC algorithm has been used by Ehlers (2011) to determine how efficient a production

function is in describing real life data. MCMC algorithms have been used by O'Donnel,

Shumway, and Ball (1999) for determining input demand functions. Finally, a Metropolis-

Hasting algorithm has also been used by Vidal, Iglesias, and Galea (2007) in observing the

influence of measurement errors in models. Thus it seems natural to apply an MCMC algorithm

to this problem, specifically the Metropolis-Hasting. With regards to the MH algorithm for this

problem, it should start with θ^0 then for j=1,2......

1. Generate$(\theta^{j^*}) \sim J_j(|\theta^{j-1})$ where $J_j(|\theta^{j-1})$ are draws from normal distributions with

means at θ^{j-1} and the standard deviation being relatively small. Since the jumping distribution is

from a normal distribution they are canceled out in the second step of the calculation.

2. Compute

$$r_j = \frac{f_x(x^{obs}|\theta^j)f_\lambda(\lambda^{obs}|\theta^j)g(\theta^j)}{f_x(x^{obs}|\theta^{j-1})f_\lambda(\lambda^{obs}|\theta^{j-1})g(\theta^{j-1})}$$

and generate

$$W_j \sim Bernoulli(min(1,r_j))$$

3. Take

$$(\theta^j) = W_j (\theta^j) + (1 - W_j)(\theta^{j-1})$$

The important point is that this algorithm generates a point to jump to, then either accepts

the point and jumps to it, or rejects it and stays at its current point. In most cases since the

number of parameters to consider are large, the algorithm tends to reject most points unless the

point proposed is relatively close to the current point. This algorithm was coded using R, how it

was coded can be found in the Appendix. It is important that the prior is not misspecified as

explained in a subsequent section because, with this limited data set, the prior can influence the

posterior.

1.6 Preliminary view of the Data

The data used is total national acreage planted obtained from the NASS from 2007-2013 for the crops of corn, soybeans, and wheat. The total acreage includes the prevented acreage (acreage that would have been planted, but were prevented most likely due to natural disaster).

Looking at the data from Table 1.1 there are a few interesting things to note, mostly looking from 2011-2013, where commodity prices vary dramatically. One key point is the fact that prevented acres seem to rise from year to year, except a drop in 2012, in which there was a drought. This makes sense because that is the year when commodity prices for corn and soybeans shot up by a significant margin. Data surrounding cost and revenue were obtained from USDA-ERS over the same time span. The crops cost were measured in per acre total operating costs, not including the total allocated overhead that the ERS reports. The data are shown in Table 1.2.

It can be seen in Table 1.2, with the exception of wheat, per acre costs are increasing over time. Revenue was measured by the product of expected price and yield. Expected yield is given by a linearization of actual yields, $Y_t = \alpha + \beta t$. This was done due to the fact that unexpected disasters like droughts could impact yields in a way farmers could not plan for. The data used in the linearization were actual average yields reported by the NASS from 1980-2013. The form of the linearization estimated in bu/acres is:

$CornExpectYield_t = 97.57 + 1.73t$

$SoyExpectYield_t = 27.35 + 0.59t$

$WheatExpectYield_t = 28.99 + 0.34t$

Here t is each year and goes from 1 to 34. As can be seen, by linearizing the yields to form expected yields has basically made the yields close to constant from 2009 to 2013.

Expected price is given by the approved projected price reported by USDA Risk Management Agency as shown in Table 1.3. This table again shows the jump in commodity prices in 2011 and subsequent years. The aforementioned tables and linearized yields, as well as the general price index, are all that is needed to simulate the posterior distribution in order to estimate this supply model. Graphically the impacts of the observed expected profitability of the crop and the observed acres that were intended to be planted are shown in Figures 1.1 and 1.2.

The general trend for Figure 1.1 shows a dip initially and then a spike occurs in 2011. Wheat for the most part is not influenced as much by the 2011 shift in commodity prices. Looking at Figure 2 gives the planned acres to plant (observed acreage plus prevented acres).

From the graphs, it appears that the total aggregate acres the farmers intend to plant match with the expected profitability, at least with respect to corn. However, there are certain trends that do not match up. The most obvious is the decrease in wheat, as expected profit clearly has an upward trend, as well as the flatness of the line in the amount of soybeans planted. It is clear competition between the crops for land usage is needed to better understand the shifts in the acreage allocated to specific crops. Thus simulations are needed to find this interaction by estimating unknown parameters generated by the PMP method.

1.7 Results of the Simulated Aggregate model

For the national crop supply model, the posterior distribution was sampled using four different starting points for θ^0, allowing the Metropolis Hasting Algorithm to iterate for five hundred thousand iterations. Since the number of parameters are relatively small, it is simple to

do such a robust check on the sampling. The starting point θ^0 was found by fitting Howitt's

(1995) initial technique in calibrating to one of the seven years between 2007 and 2014. A guess

for own price elasticity was drawn between 0 and 1 for each crop and used to calibrate the

unknown quadratic as a diagonal matrix for a given year randomly selected between 2007 and

2014. This was all done to produce varying θ^0s, while at the same time making sure that the

point is in fact a feasible result.

Thus we have a starting point where for a randomly selected year $\lambda(\theta^0) = mean(\lambda^{obs})$,

$x(\theta^0) = mean(x^{obs})$ and $\eta(\theta^0)$ is at some value in the ball park of its distribution η_{guess}. At

this starting point, the only major error that the MCMC algorithm is trying to correct for is what

the believed prior is and the inaccuracy in its prediction of the other years. Therefore the

algorithm is selecting θ such that $g(\theta)$, $f_x(x^{obs}|\theta^j)$ and $f_\lambda(\lambda^{obs}|\theta^j)$ in a hill climbing algorithm

trying to find the θ that best satisfies these distributions.

The parameters for θ were examined over its inverted parameters for the quadratic

parameters, i.e. $U_{ii}^* = \frac{1}{U_{ii}}$. This was done mainly because of the fact that Q ultimately gets

inverted, so small values of θ would account for large elements of Q^{-1}, which in turn would

significantly impact the optimal amount of acreage chosen by the farmer. Thus θ^* is used so that

the jumping distribution used in the algorithm would not cause large shifts in the acreage with

each new point.

Describing the posterior distribution based off the likelihood function and prior

distribution is extremely difficult. However, in simulating the posterior distribution there is a

way to view what this distribution looks like. Figures 1.3 and 1.4 give the trace and plot for all

six parameters that make up θ. The major take way from these plots are that the posterior

distributions are single peaked for each variable.

The estimated median of the resulting posterior distribution formed from this MCMC

algorithm for the unknown θ^* is shown in Table 1.4. These results show a relatively tight range

on all the parameters for θ^*. The fact that $Q = U'U$ implies by construction that unknown

quadratic cost term is such that the marginal cost of producing a specific crop is increasing with

respect to its own crop.

Next one must examine how well the optimum acreage of the model matches the data it is

based upon. Table 1.5 examines the distributions of the optimum acreage based on the posterior

distribution of θ^*. The measurement is both at the median of the distribution for the simulation

and 95% credible interval is shown in parentheses. The predicted values are from the 95%

credible interval from a sample distribution of four chains. As observed above, the resulting

prediction based on the optimization modeling have intervals which for all years encompass the

observed acreage and in many cases the observed value is relatively close to the median of the

distribution. Of note, the only year where the model has trouble predicting is 2011, the year a

major shift in expected profits occurred. Other than that instance, these results reflect the benefit

of employing a programming method creating an optimization model that can reproduce

observations. The elasticity of the posterior can also be seen by looking at the median and 95%

credible interval (parenthetic) from a sample distribution. By examining the elasticity matrix for

a single year (2013) the median of the elasticity is shown in Table 1.6.

Examining the literature, Hendricks, Smith, and Sumner (2014) measure long run own

price and cross price elasticities between corn and soybeans: their estimates suggest an own price

for corn at 0.29 and soybeans at 0.26. In the short run, these elasticities are 0.40 for corn and .36

for soy. Their results have short run elasticity being greater than long run elasticity, which seems like an odd result. However, the elasticities their model found could still be seen as a relative range the own price elasticity could be. The resulting elasticity for this chapter's model should represent the short run elasticity, so the overall results from Table 1.6 are showing a more inelastic supply elasticity. A regional model might better replicate the resulting elasticities shown by Hendricks, Smith, and Sumner, again only needing a limited amount of information. Looking at the elasticities generally, the own price elasticity is inelastic and generally soybeans and corn are similar in terms of own price elasticity, while wheat appears the most inelastic. It is also interesting to note that the strongest cross price elasticity lies between corn and soybeans, and the cross price elasticity for wheat tends to be fairly weak, if it does exist.

1.8 Convergence Test

The robustness of the model can be checked through the Brooks-Gelman-Rubin (BGR) test. This was done for four sample models for over five hundred thousand iterations. The general formula for this test for an individual element is as follows. Let m be the number of samples, and n the number of iterations in sample and let ψ be a sample element within θ so the simulated draws are $\psi_{i,j}$ for i=1,..,n and j=1,..,m. Then the between and within sequence variances are calculated as:

$$B = \frac{n}{m-1}\sum_{j=1}^{m}\left(\psi_{*j} - \bar{\psi}_{**}\right)^2$$

$$W = \frac{1}{m}\sum_{j=1}^{m}\left(\frac{1}{n-1}\sum_{i=1}^{m}\left(\psi_{i,j} - \bar{\psi}_{*j}\right)^2\right)$$

With that the variance can be derived as :

$$\widehat{var} = \frac{n-1}{n}W + \frac{1}{n}B$$

From here the convergence is then measured by

$$R = \sqrt{\frac{\widehat{var}}{W}}$$

This will converge to 1 as n→∞.This will allow us to determine how many iterations are suitable for the sampling to converge. However, as this is a test for a multivariate model, the added complication to this test will be dealt with by using an R package, called coda created by Plummer(2015). This will allow for the measurement of this variability at a multivariate level.

The rule of thumb for this test is that the chains have converged if R is at 1.1 or lower. Looking at the point estimates in Table 1.7 they are all under 1.1. Thus it can be said there is convergence within five hundred thousand iterations.

1.9 Alternative Prior

In addition, an alternative prior was considered as a robustness check. The prior assumed an elasticity of 1 across all crops. This would change the model in that for U_{ii} the prior assumed is g(U_{11})~$N(5785, (700)^2)$, g(U_{22})~$N(6566, (700)^2)$, and f(U_{33})~$N(9597, (700)^2)$. Convergence for this model was found using a BGR test. Examining the results, the overall convergence for the parameters were found to be the shown in Table 1.8. The most interesting part to this posterior distribution is that U_{11}, which here represents the unknown quadratic parameter for the planting of corn, is almost exactly the value of the previous posterior distribution despite the change in prior.

Looking at the model's result in Table 1.9, it is clear that while this model with an alternative prior tries to calibrate itself to the observations, it cannot do it as well with the

misspecified prior. The best match up in Table 1.9 is with acres of corn, and that is probably due to the fact it reverted to the posterior density assumed in the correct model. The median for the own price elasticity was found for this model in 2014 and was (.35, .465, .47) for corn, soy, and wheat. So while this prior might assume the crops should be more elastic, data suggests otherwise. It is interesting to note how strong the data is suggesting that corn should be near .35, and the overall accuracy of the model in matching corn observations. This might suggest why it is easier for the model to calibrate to shifts in corn price, than to soy or wheat, as the data might support this conclusion.

1.10 Forecasting

For prediction I used the year 2014. It is easy enough to project the expected yields and an expected cost for 2014. The expected costs are found from trending the cost data and projecting it forward one year. Similarly, expected yields can be projected forward one year and expected profits per acre for each crop can be found. The only thing missing in this forecast is the total land being allocated in 2014. For the most part, total land barely changes between years, but it does in small part as the shadow price of land increases. Taking the shares of each crop allocated in each region in the previous year (2013), an expected shadow value of land can be found. Using that expected value for the shadow land, the change in overall value is examined between 2013 and 2014, and with that shift alongside the trend for how the shadow price of land influences total land, an expected value for total land in 2014 can be found. All the values are measured in acres both planted and prevented planted for the year. These predictions are shown in Table 1.10.

As can be seen in Table 1.10, this aggregate model has modes fairly close to those observed for Corn and Wheat. However, the prediction is slightly higher than the mode for soybeans. This will be corrected by employing more information in a regional model. Ultimately this Bayesian technique does an impressive job at forming a posterior distribution that can actually reflect the observations, while at the same time forming reasonable estimates for the elasticities and produce reasonable forecasts.

1.11 Conclusion

The work done in this chapter has shown how to employ an MCMC algorithm to determine the set of parameters that are consistent with the beliefs of what a national crop supply model should be, in the context of a PMP model. The general results of the model show promise in that they reflect what is observed and appear fairly consistent with the literature. Improvements to the model, specifically to a more informed prior, could be made that allow for credible intervals without a tremendous amount of sampling needed.

Chapter Two expands this national model into a regional model, using regional data published by USDA-ERS. The only real hurdle to this endeavor is to figure out how to disaggregate the prior from the national level to the regional level. The results of the model show how this basic programming model can, with some brute force, fit the non-linear problem, while keeping consistent to economic theory that it is embedded in.

1.12 Reference

Arfini, F., Donati, M., Grossi L. and Pariss, Q. Revenue and Cost Functions in PMP: a Methodological Integration for a Territorial Analysis of CAP. Paper presentation at the

107th EAAE Seminar "Modelling of Agricultural and Rural Development Policies". Sevilla, Spain, January 29th -February 1st, 2008.

Buysse, J. Fernagut, B., Harmignie, O., Henry de Frahan, Lauwers, L. Polome, P., Van Huylenbroeck, and Van Meensel. Farm-Based Modelling of the EU Sugar Reform: Impact on Belgian Sugar Beet Suppliers. European Review of Agricultural Economics 34(1): 21-52, 2007.

Carrol, R., Ruppert, D. and Stefanski, L. Measurement Error in Nonlinear Models. Suffolk, UK: Chapman and Hall, 1995.

Chambers, R., and Just R. Estimating multi-output technologies. American Journal of Agricultural Economics 71:980-995. 1989.

Ehlers, R. Comparison of Bayesian models for production efficiency. Journal of Applied Statistics. Vol. 38, No.11, Nov 2011.

Gocht, A. Methods in Economic Farm Modelling. PhD dissertation, Department of Economic and Agricultural Policy, University of Bronn. Bronn, Germany. 2009.

Gleman, A., Carlin, J., Stern, H., and Rubin, D. Bayesian Data Analysis: Second Edition. Chapman and Hall, 2004.

Heckelei, T. and Britz, W. Models based on Positive Mathematical Programming: State of the Art and Further Extension. In Arfini, F. (eds), Modelling Agricultural Policies: State of the Art and New Challenges. Proceedings of the 89th European seminar of the European Association of Agricultural Economics. Parma, Italy: University of Parma, 48-73. 2005.

Heckelei, T. Wolff, H. Estimation of constrained optimizations models for agricultural supply analysis based on generalized maximum entropy. European Review of Agricultural Economics Vol 30(1) pp.27-50. 2003.

Hendricks, N., Smith, A., and Sumner, D. Crop Supply Dynamics and the Illusion of Partial Adjustment. Amer. J. Agr. Econ 96 (5): 1469-1491, May 2014.

Holt, M.T. A Linear Approximate Acreage Allocation Model. Journal of Agricultural and Resource Economics 24(2):383-397. 1999.

Howitt, R. Positive Mathematical Programming. Amer. J. Agr. Econ. 77: 329-342, May 1995.

Jansson T. and Heckelei T. Estimating a Primal Model of Regional Crop Supply in the European Union. Journal of Agricultural Economics, Vol.62, No.1, 2011.

Kasnellopoulos, A. Berentsen, P., Heckelei, T., Van Ittersum, M. and Oude Lansink. Assessing the Forecasting Performance of a Generic Bio-Economic Farm Model Calibrated with Two Different PMP Variants. Journal Of Agricultural Economics 52(7):333-340. 2010.

Merel, P.R. and Bucaram, S. Exact Calibration of Programing Models of Agricultural Supply against Exogenous Sets of Supply Elasticities. European Review of Agriculture Economics 37:395-418, 2010.

O'Donnell, C., Shumway, C., and Ball, V. Input Demands and Inefficency in U.S. Agriculture. Amer. J. Agr. Econ 81(865-880), Nov 1999.

Paris, Q. and Howitt, R. An Analysis of Ill-Posed Production Problem using Maximum Entropy. Amer. J. Agr. Econ. 80: 124-128, February 1998.

Plummer, Martyn. (2015) coda: Output Analysis and Diagnostics for MCMC. R package version 0.18-1

Vidal, I., Iglesias, P., and Galea, M. Influential Observations in the Functional Measurement Errror Model. Journal of Applied Statistics. Vol. 34, No.10, Dec 2007.
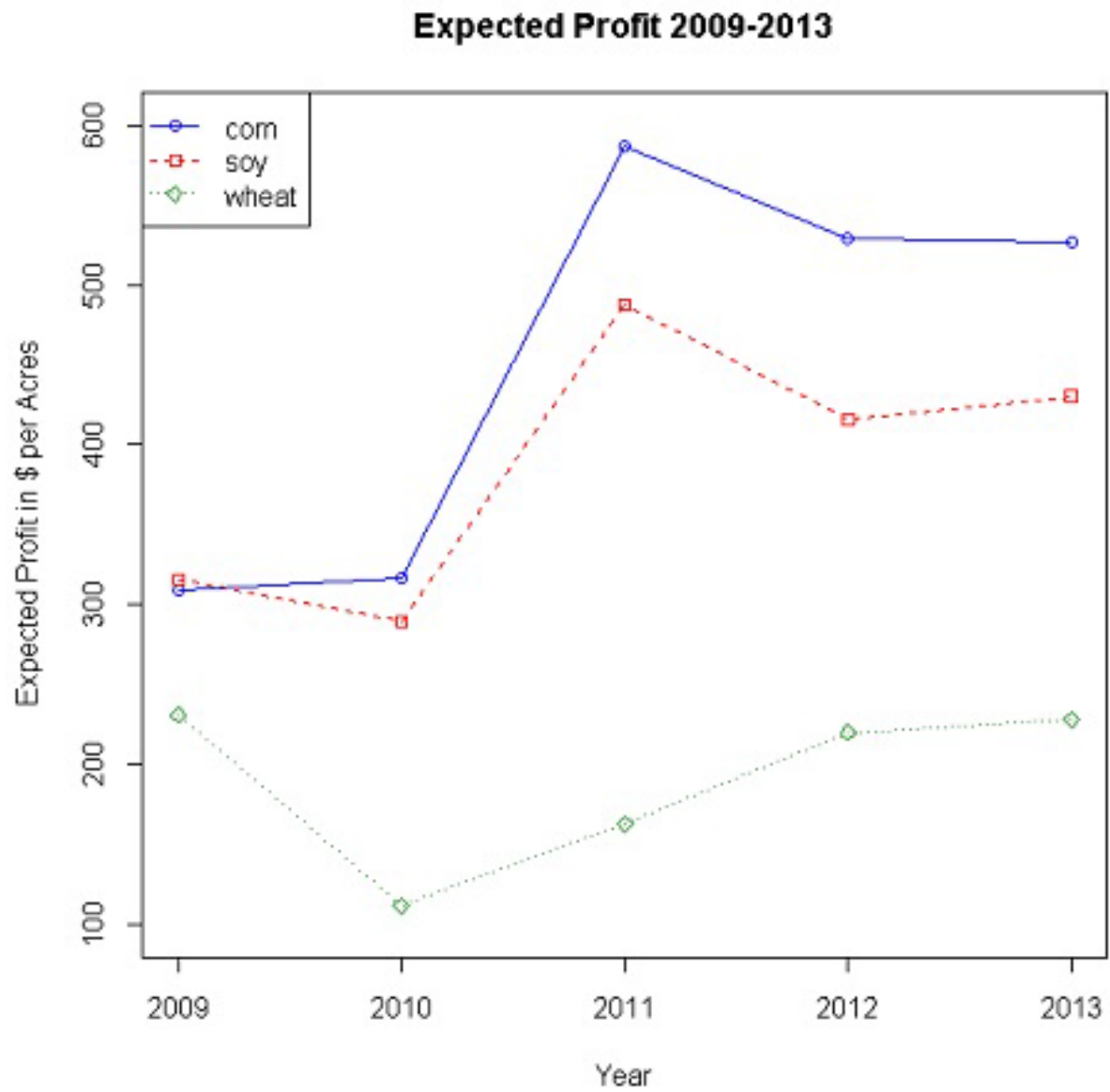
Figure 1.1 Expected Profit 2009-2013
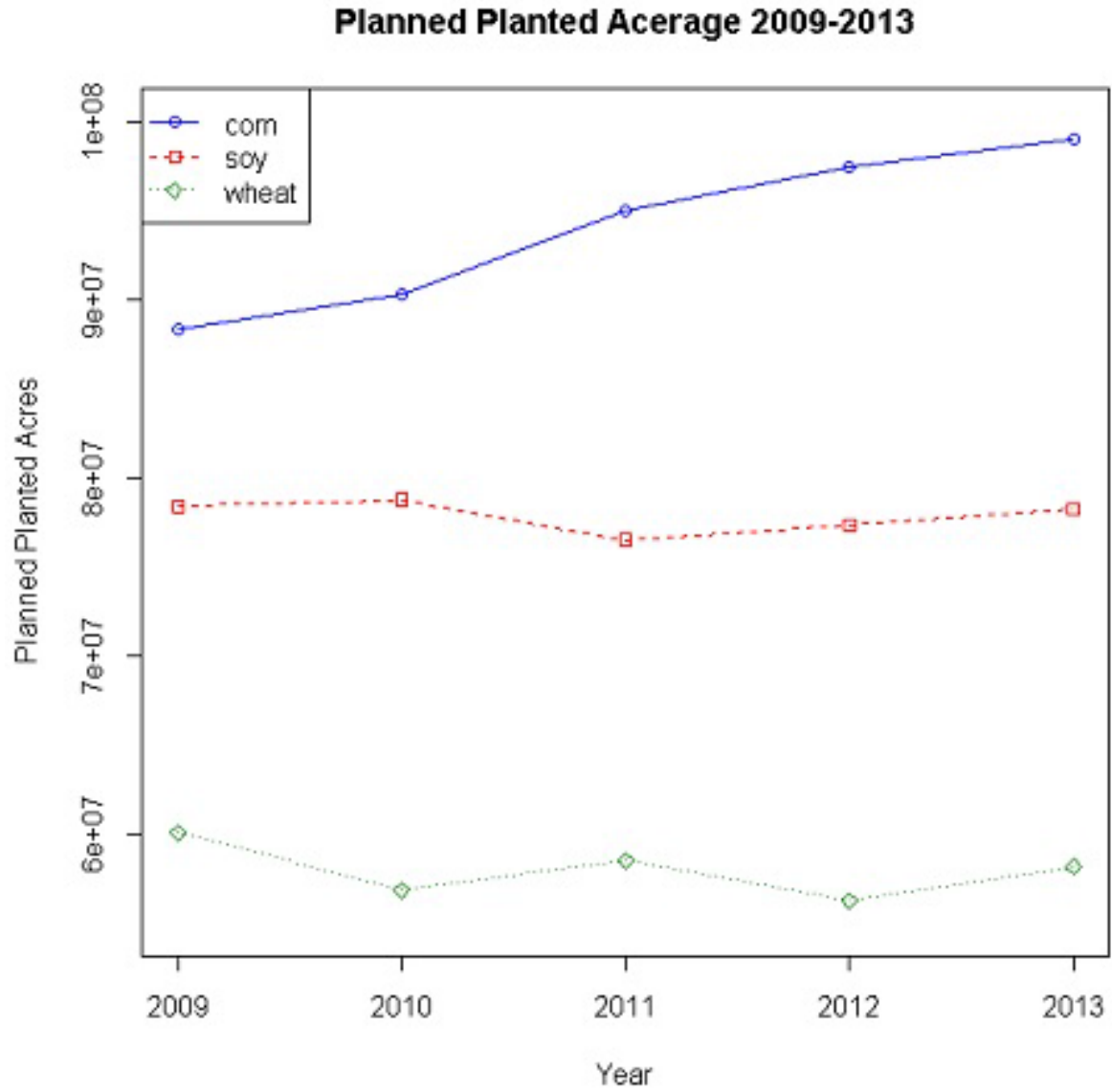
Figure 1.2 Planned Planted Acreage 2009-2013
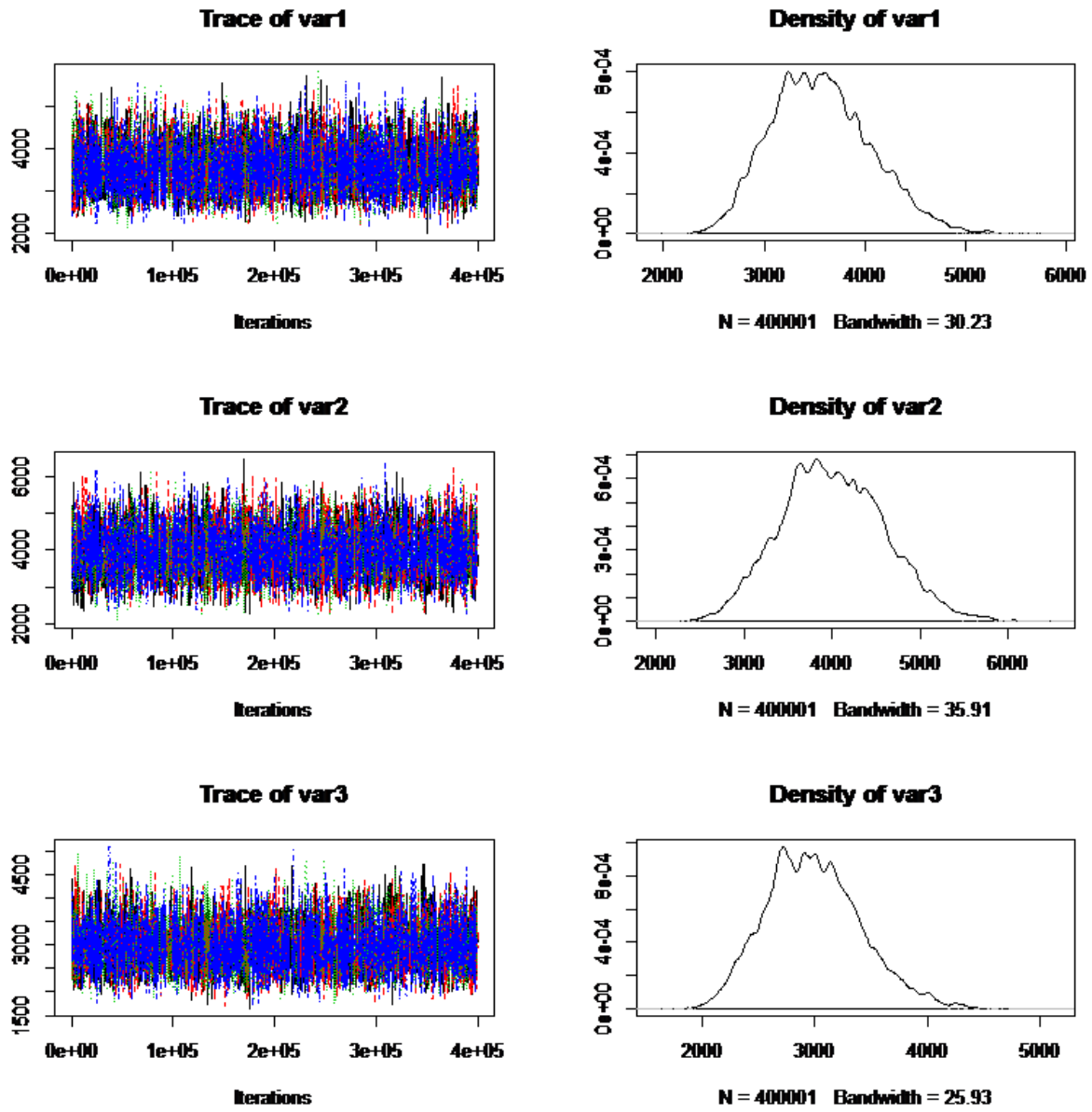
Figure 1.3 Posterior Density and Trace Plot of National Model Part 1; Var1=U_{11}, Var2=U_{22}, Var3=U_{33}
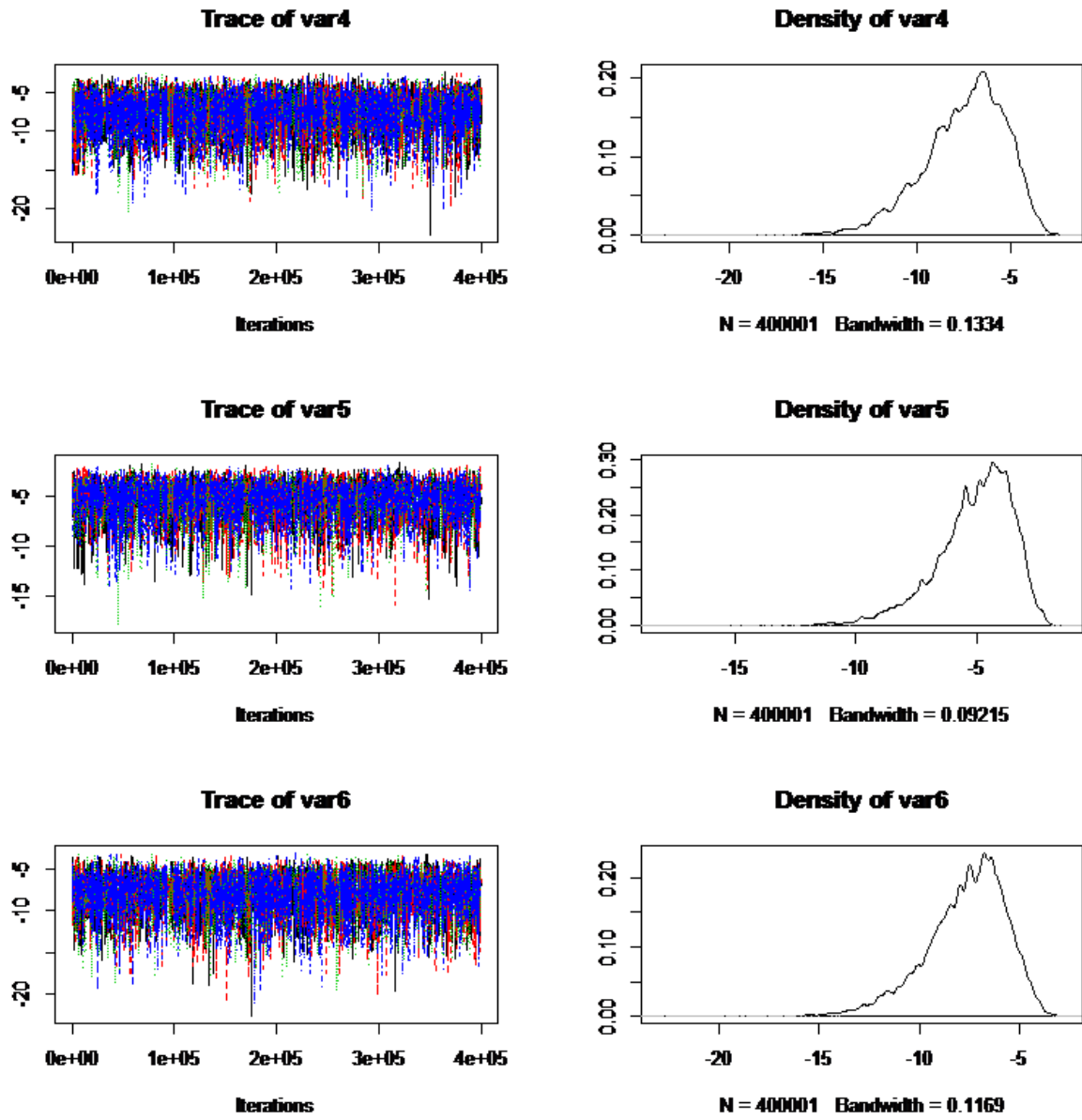
Figure 1.4 Posterior Density and Trace Plot of National Model Part 2; Var4=h_1, Var5=h_2, Var6=h_3

Table 1.1: Planted and Prevented Planted Acres in Millions						
Years	Corn Acres		Soy Acres		Wheat Acres	
	Planted	Prevented	Planted	Prevented	Planted	Prevented
2013	95.365	3.617	76.533	1.704	56.156	2.013
2012	97.155	0.262	77.198	0.160	55.666	0.587
2011	91.936	3.013	75.046	1.447	54.409	4.117
2010	88.192	2.102	77.404	1.347	53.593	3.258
2009	86.382	1.879	77.451	0.933	59.168	0.917
Source: National Agricultural Statistics Service, USD						

Table 1.2: Average Cost in $ per acre			
Year	Corn	Soy	Wheat
2013	$355.98	$180.36	$128.08
2012	$349.59	$172.29	$126.72
2011	$332.33	$136.87	$121.89
2010	$286.41	$131.89	$102.78
2009	$295.01	$130.49	$112.92
Source: ERS			

Table 1.3: Expected Price of Crop in $ per Bushel			
Year	Corn	Soy	Wheat
2013	$5.65	$12.87	$8.78
2012	$5.68	$12.55	$8.62
2011	$6.01	$13.49	$7.14
2010	$3.99	$9.23	$5.42
2009	$4.04	$9.90	$8.77
Source: USDA Risk Management Agency			

Table 1.4: Posterior distribution of θ at the Median & 95% Credible Interval			
Quadratic portion of $\theta^* \left(\frac{Acres}{\sqrt{\$}} \right)$	U_{11}	U_{22}	U_{33}
Estimated Median	3536	3981	2969
(95% Credible Interval)	(2718, 4615)	(2915, 5215)	(2237, 3913)
Linear portion of $\theta^* \left(\frac{\$}{Acres} \right)$	h_1	h_2	h_3
Estimated Median	-7.13	-4.83	-7.38
(95% Credible Interval)	(-12.33, -4.00)	(-9.09, -2.73)	(-12.39, -4.53)

Year	Corn Acres	Corn Acres Obs.	Soy Acres	Soy Acres Obs.	Wheat Acres	Wheat Acres Obs.
	Table 1.5: Distribution of Predicted Crop Acreages 2007-2013 (in Millions) at the Median and 95% Credible Interval					
2009	88.41	88.26	77.05	78.38	61.29	60.08
	(84.45,92.45)		(73.11,80.87)		(57.72,64.95)	
2010	90.49	90.29	77.54	78.75	57.83	56.85
	(86.78,94.38)		(73.80,81.21)		(54.80,60.81)	
2011	96.97	94.94	79.31	76.49	53.74	58.52
	(92.93,101.1)		(75.36,83.22)		(50.16,57.07)	
2012	96.12	97.41	77.07	77.35	57.85	56.25
	(92.20,100.1)		(73.21,80.75)		(54.80,60.82)	
2013	97.69	98.98	78.6	78.23	59.11	58.16
	(93.63,101.8)		(74.53,82.46)		(55.87,62.25)	
Measured in millions of acreage, and includes prevented acres in both predicted and observed						

	Corn Price	Soy Price	Wheat Price
Table 1.6: 2013 Supply Elasticity based on the posterior distribution			
Corn Acres	0.255	-0.185	-0.1301
	(0.170,0.375)	(-0.297,-0.106)	(-0.223,-0.071)
Soy Acres	-0.1189	0.237	-0.1208
	(-0.197,-0.067)	(0.157,0.335)	(-0.202,-0.066)
Wheat Acres	-0.0513	-0.074	0.171
	(-0.091,-0.027)	(-0.129,-0.04)	(0.112,0.246)

Table 1.7: Measurement of $\hat{R}$ for all elements in θ^*							
Point estimate	U_{11}	U_{22}	U_{33}	h_1	h_2	h_3	Multivariate psrf
Measurement of $\hat{R}$	1.00	1.00	1.00	1.00	1.00	1.00	1.01

Table 1.8: Posterior distribution of θ at the Median & 95% Credible Interval (Misspecified)			
Quadratic portion of $\theta^* \left(\frac{Acres}{\sqrt{\$}} \right)$	U_{11}	U_{22}	U_{33}
Estimated Median	3544	4553	8050
(95% Credible Interval)	(2418, 5157)	(3242, 6519)	(5649, 10020)
Linear portion of $\theta^* \left(\frac{\$}{Acres} \right)$	h_1	h_2	h_3
Estimated Median	-7.08	-3.63	-1.67
(95% Credible Interval)	(-15.72, -3.05)	(-7.28, -1.74)	(-2.63, -1.15)

Table 1.9: Distribution of Predicted Crop Acreages (in Millions) 2007-2013 at the Median and 95% Credible Interval (Misspecified)					
Year	Corn Acres	Corn Acres Obs.	Soy Acres	Soy Acres Obs.	Wheat Acres	Wheat Acres Obs.
2009	84.56 (79.38,89.53)	88.26	70.95 (66.63,75.50)	78.38	71.17 (66.85,75.58)	60.08
2010	89.04 (85.06,92.89)	90.29	75.55 (71.96,79.42)	78.75	61.23 (58.18,64.29)	56.85
2011	100.33 (95.43,105.4)	94.94	84.67 (80.16,89.59)	76.49	44.85 (40.30,49.50)	58.52
2012	96.86 (92.76,101.0)	97.41	77.14 (73.29,81.02)	77.35	56.93 (53.78,60.12)	56.25
2013	99.26 (95.06,103.3)	98.98	79.56 (75.43,83.70)	78.23	56.55 (53.16,60.01)	58.16
Measured in millions of acreage, and includes prevented acres in both predicted and observed						

Table 1.10: Prediction for Crop Acreage for 2014 (in Millions), Median & 95% Cred. Interval		
2014	Predicted	Observed
Corn	89.29	92.46
	(85.44,93.29)	
Soy	79.22	85.01
	(75.34,83.09)	
Wheat	57.84	58.20
	(54.81,60.88)	

CHAPTER II: ESTIMATING REGIONAL CROP SUPPLY MODEL ELASTICITIES USING PMP AND BAYESIAN ANALYSIS

2.1 Abstract

In the previous chapter, a national model was derived using a mix of Positive Mathematical Programing and Bayesian Analysis. The model generates a distribution for the calibrating parameters that could reproduce the observe acreage at the optimal acres for a farmer to plant, and could forecast the decision of farmers in the US for future years. The model in this chapter expands upon the previous national model by estimating regional crop supply models. This is done while overcoming a greater uncertainty with respect to the supply elasticity at a regional level, by using the information of the supply elasticity at a national level and disaggregating it to a regional level. Modeling is done in four regions that make up 85% of production in corn, soy, and wheat in the US. Not only is the regional model formed capable of forecasting future land allocation at the regional level, but it does so in a more accurate fashion than the national model. This is due to the fact that a regional model can take into account the specialization that can occur within each region. Moreover, this model does what no other model has done: provide a range of possible regional supply elasticities for corn, soy, and wheat in the US.

2.1 Introduction

Being able to properly model the supply of US crops has been a difficult task to master. As shown in the previous chapter, there are many ways for this modeling to be done. However, to get around some of the more difficult aspects of structural changes that occur in this modeling, we utilized both techniques used in Howett's (1995) Positive Mathematical Programing and in

Bayesian Analysis. In the previous chapter, we successfully estimated a national model that could, for the most part, be used to forecast the future supply of corn, soy, and wheat for the US. There are, however, difficulties in that model that this chapter remedies. In this chapter the model and its forecasting ability are improved by utilizing these techniques at a regional level.

The ERS (Economic Research Service for the USDA) has attempted in their analyses to take into account various agricultural regions that exist for the US. As of 1995, they divide the US into 9 distinct regions. These regions are the Heartland, Northern Crescent, Northern Great Plains, Basin Range, Fruitful Rim, Prairie Gateway, Mississippi Portal, Southern Seaboard, and Eastern Uplands. In the present examination of estimating a regional model I restricted my attention to the most important regions that produce corn, soy, and wheat. This ultimately narrows to four regions: the Northern Crescent, including Wisconsin and Michigan; The Heartland, mainly Iowa, Illinois, Ohio, Indian, and Missouri; Northern Great Plains, North and South Dakota and parts of Nebraska; Prairie Gateway, Kansas, part of Nebraska, New Mexico, and Colorado, as well as most of Oklahoma and Texas. These four regions combined produce about 85% of the US's production of corn, soy, and wheat.

There are many difficulties in applying the technique used at the national level to a regional level. The largest caveat is that of forming a prior on the supply elasticity at a regional level. However, by employing what little knowledge is known about regional crop supply elasticities, that is, the fact that regional supply elasticities aggregate to a national elasticity, a model for determining the supply of crops at a regional level can be found. This is done by layering the prior. This is accomplished by adding in a prior with respect to the portion each region contributes to the slope of the national supply curve. Having accomplished this, the analysis at a regional level is similar to that of the national model. The results of this chapter

shows that with the additional data at a regional level, a more accurate model is formed by taking into account the specializations that can occur within the regions. In addition, the model can accurately forecast 2014 acreage with no additional information except expected prices. Finally, this work provides a view of the supply elasticity at a regional level which is lacking in previous literature.

2.3 Literature Review of Regional Models

The division of regions as classified by the ERS for the US was done in 1995. There has been no apparent research on regional supply elasticities. This might be due to the fact that this new reclassification of regions only occurred in 1995. Methods that require more data might be ill suited to form an accurate econometric model for these regions. In addition, there has not been any apparent application of Howett's PMP method to calibrate regional models. This makes the results of this regional model's supply elasticities difficult to compare with other literature as none exist. There still, however, exist studies which have used regional data classified by the ERS in some form or another, for other purposes.

One study done by Fernandez-Cornejo and Wechsler (2012) tried to determine the impact of using Bt corn (genetically engineered crop with Bacillus thuringiensis) on the profits and use of insecticides on farms. The overall regression suggested that the adoption of Bt corn did not impact the use of insecticides. The study, however, was limited to 2005 data; so the result is questionable. The interesting use of regional information came from the fact the study wanted to isolate farmer decisions in the Heartland compared to national data. This focus was done primarily because the opportunity cost of pest infestations were higher in the Heartland than in other regions. The Heartland effect was ultimately a dummy variable that normally accounted for

higher yields in the impact study. This study is an example of how regional data can be used to try and isolate regional effects on the farmers' decision making.

One reason to try and pin down the various regional effects is illustrated by Goodwin and Mishra (2006). They wanted to determine if US farmer program benefits were in anyway distorting production. The focus of this study was on commercial farming, and so their analysis was confined to survey data from the Heartland region. In many ways it makes sense to focus on the Heartland, due to its heavy production in agriculture. However, there needs to be a better understanding of the Heartland's production, and how it differs from other regions, in order to better understand how this would impact the US. A different study makes a better case for how regions interact with national models.

Research by Nehring, Fernandez-Cornejo, and Banker (2005) attempt to determine off-farm labor and earnings for corn and soybean farmers. Their findings suggest that smaller corn and soybeans farmers are increasing their off farm earnings due to the competitiveness of larger farmers. What is interesting in terms of relating this to regional modeling is that their study included only corn and soy, so that they only had to examine the regions of the Heartland, Northern Great Plains, Northern Crescent, and Prairie Gateway. The study could have included wheat, as it is mostly produced in the Northern Great Plains and Prairie Gateway, but the authors chose to aggregate it as 'other crops' in their study.

The examination of regions becomes increasingly important when considering policy impacts. A study done by Goodwin, Vandeveer, and Deal (2004) examined the impacts of crop insurance on the acreage response for corn, soybeans, wheat and barley. Their findings were that the increase in acreage due to crop insurance was modest at best, and mostly only in wheat. The study, however, only examined the Heartland and the Northern Great Plains. Considering the

larger model, it is also important to be able to consider whether or not the Northern Crescent would behave differently from the Heartland, as the total acreage in that region is small and might be more responsive to change. It is also interesting that the Prairie Gateway region was not considered in the study. The Prairie Gateway is similar to the Northern Great Plains in terms of a focus on wheat, but might behave differently in terms of policy impact.

The need for studying regional models was demonstrated by Henderson (2008) who examined the agricultural boom that started in 2007. This boom occurred primarily from a combination of ethanol demand and weaker global production due to droughts. This article examined the capitalized returns to corn for corn production in the four regions of Heartland, Northern Crescent, Northern Great Plains, and Prairie Gateway, finding that the Heartland had far and beyond the largest returns, being twice that of Northern Great Plains and four times that to Northern Crescent and Prairie Gateway. Thus when considering boom or policy impacts, one needs to consider how the different regions are impacted.

In a more recent article Garnache and Merel (2015) attempted to see if it is possible to disaggregate the supply response found with PMP modeling. Their model extends the PMP from a fixed land constraint to multiple constraints using a fixed portions model that allows for strict concavity on the production function. Their method then builds an algorithm to disaggregate the supply elasticity by using a aggregated elasticity alongside regional output shares and additional regional calibration criteria to disaggregate the elasticity. This technique was possible even with a limited amount of information with regards to acreage, yields, prices, and constraints. The algorithm posed in their model to disaggregate the elasticity is similar to the method used in this chapter in forming the regional prior.

All of these studies highlight the difference between the various regions in the US in terms of production and price response. It is curious there has been no attempt to study the supply elasticity at a regional level. It is obvious that farmer decision making in the Heartland is different from decision making in the Northern Crescent, or Northern Great Plains, or in the Prairie Gateway. Hence a better regional model is needed, and with a programming model it might be possible, even if the data is limited.

2.4 Regional Crop Supply Model

The model developed in Chapter 1, that show how farmers allocate their land on a national level could also be used in describing land allocation at regional levels. To achieve this type of modeling we need information to calibrate the model to observed data. Thus the nonlinear programming problem that calibrates the regional model is similar to the national model and the general form is again similar to the optimization that Jansson and Heckelie(2011) used:

$$\max_{x_{tr}} x'_{tr}[p_t Y_{tr} - cost_{tr}] - q_t x'_{tr}\left[h_r + \frac{1}{2}l_{tr}Q_r x_{tr}\right]$$

$$\text{s.t. } R_t x_{tr} = v_{tr}$$

(1)

Here r represents the region being considered, i.e., the Heartland, Northern Crescent, Northern Great Plains, and Prairie Gateway. t represents the years from 2009-2013. Again there is x_{tr}, a vector of acreages for J crops, Y_{tr} is a J×J matrix of yields, p_t is a J vector of national expected prices. National expected prices were used in part because in the next chapter we want one market to sell the various crops, and not several regional markets. In addition, looking at expected prices from USDA Risk Management Agency, the expected price does not significantly

change between the regions over the five year time span. Additional variables used are $cost_{tr}$ which is a J vector of average cost per acre , q_t is a J vector of price index, h_r is a J vector of the linear behavior parameters that are to be estimated for the region, l_{tr} is the land availability index, used to take into account any shifts in the total available land between years. Q_r is a J×J diagonal matrix of own crop effects to be estimated for the region. Here Q_r is diagonal due to the fact that the additional complexity to this variable would make calculating a prior for this based on a national slope for the model extremely difficult. R_t is a J vector of ones and v_{tr} is total land available. The first order and second-order conditions are then:

$$p_t Y_{tr} - cost_{tr} - q_t h_r - q_t l_{tr} Q_r x_{tr} - R_t \lambda_{tr} = 0 \tag{2}$$

$$R_t x_{tr} = v_{tr} \tag{3}$$

$$Q_r = U_r' U_r \tag{4}$$

Since the optimization model is the same as for the national model, the general solutions for finding optimal acreage is also the same. It is easy to form functions for acreage, shadow value of land, and elasticity at a regional level. The main difficulty in this model stems from the lack of information about regional elasticities. There is no literature that forms a basis for what the elasticity should be in any of the four regions being considered for any of the three crops. There is, however, a relationship between national and regional elasticities which could be exploited. In examining $\frac{\partial x_{tr}^*}{\partial p_t}$ for this problem it is found that:

$$\frac{\partial x_{tr}^*}{\partial p_t} = E_{tr}^{-1} Y_{tr} - E_{tr}^{-1} R_t [R_t E_{tr}^{-1} R_t']^{-1} R_t E_{tr}^{-1} Y_t \tag{5}$$

Where $E_{tr} = q_t l_{tr} Q_r$ and Y_{tr} is a diagonal matrix of the yields for the crops in the region. However, the information that can be used in this model is for a national slope, referred to as

$\frac{\partial x_{tn}^*}{\partial p_t}$. The only relationship that exists between national and regional is that the regional slopes

should aggregate to a national slope for the supply of the various crops. As a result, the

information at a national level needs to be disaggregated to a regional level. However, how is

information of the national slope supposed to be disaggregated to a regional slope? Looking at

the problem, one condition that links the regional optimization to the national model is the fact

that all regional models must satisfy the following condition:

$$\frac{\partial x_{tn}^*}{\partial p_t} = \sum_{r=1}^{r=9} \frac{\partial x_{tr}^*}{\partial p_t}$$

The problem here is that such a form of aggregation leaves a lot of unknowns for what

the values $\frac{\partial x_{tr}^*}{\partial p_t}$ could take for the model. Since only a subsection of the regions are used, it is

important to the model that only those four regions matter. Due to uncertainty surrounding these

variables, this condition should be constrained as an average across all years. This overall

restriction can be rewritten as:

$$\frac{\partial x_{tr}^*}{\partial p_t} = \alpha_r \frac{\partial x_{tn}^*}{\partial p_t}$$

$$\text{s.t.} \sum_{r=1}^{r=9} \alpha_r = 1$$

Here α_r is the percentage of the national price slope for the crop represented by the

region r. In doing so we now have a new unknown parameter for which a prior could be formed.

However, in considering a prior of this type, it is essential to incorporate some general belief to

ensure the model will converge. One basic belief is that the price responsiveness for the three

crops for a single acre should be the same across all acres. Thus looking at the total land acreage

for each of the four regions across the three crops as a percentage of the total land acreage

nationally gives a basis for the prior on α_r. It is thus possible to form a regional slope based on such a prior, and thus a prior for what the quadratic parameters are to be. Then a distribution could be formed for each region such that the model reflects some weak belief on a regional elasticity which can be aggregated to the national elasticity.

From the previous national model, I was able to create a posterior distribution for the national unknown parameters based on observations and a previous belief concerning elasticity. With this posterior distribution, we have a more informed belief on what the price slope should be with each crop nationally. It is this information that can be incorporated to more accurately form a posterior density for the regional parameters. By using this posterior distribution on the national slope, we can draw from it allowing the regional models posterior density to be conditional on that national slope distribution. Overall this would probably make the posterior density more diffuse, as it allows the model to take into account the uncertainty of what the national slope should be.

2.5 Bayesian estimation and Metropolis-Hasting Algorithm

The modifications to the prior are the only major changes needed in generalizing the national model to the regional model. Since a prior is based off of the aggregation of the slopes, this means that the prior has to involve all regions and cannot be done separately for each region. As a result, the overall posterior density is more complex, however, it is still easy to break down the distribution into its various components as most of the densities should be independent of other densities. First, utilizing the general Bayes rule:

$$g(\theta|z) \propto f(z|\theta)g(\theta) \tag{6}$$

In this problem the posterior density that is being found for $\theta = (U_r, h_r, \alpha_r, for\ r = 1..4)$ and the given data is the observed acreage, expected revenue or $p_t Y_{tr} = revenue_{tr}^{obs}$, the observed cost, and the shadow price of land as measured by the average expected profit per acre across all crops. So then $z = (x_r^{obs}, revenue_r^{obs}, cost_r^{obs}, \lambda_r^{obs}\ for\ r = 1..4)$. Now all that is left is to define the likelihood function and prior density.

The first prior density to consider is for α. The easiest way to form this prior, is to use a Dirichlet distribution (in order to sum to one) for all three crops, say $\prod_{i=1}^{3} dirch(\alpha_i)$ with its vector being centered for all three crops at $\alpha_i = (0.42, 0.08, 0.15, 0.18, 0.17)$. These values are the percentage of regional to national for the agricultural land for corn, soy, and wheat for Heartland, Northern Crescent, Northern Great Plains, Prairie Gateway. The last value of α is the aggregation of the five other regions not being considered. Since a change in α can have a dramatic affect on all other parameters, the variance is tightened by having the scale value set at 100. This would mean that the variance on the prior for each of the elements is about 0.01. While restrictive, this variance still allows for the model to consider alternative α's.

In addition the posterior density for the national slope is used in this problem from the previous chapter, defined as $g_{slope}\left(\frac{\partial x_n^*}{\partial p}\right)$, which is assumed given. From this, a conditional density for U can be formed.

Finding a prior density for U which is conditional on both α and the national slope $\frac{\partial x_n^*}{\partial p}$ for the three crops, is defined as $\prod_{r=1}^{4} g_{U_r}\left(U_r | \alpha_{corn}, \alpha_{soy}, \alpha_{wheat}, \frac{\partial x_n^*}{\partial p}\right)$. Since U is a diagonal matrix it is possible for find a relationship for these parameters by solving for the second order condition (equation 5) in terms of U_{iir} which gives:

$$U_{11r} = -\frac{(-2ab - 2ac - 2bc + a^2 + b^2 + c^2)}{(-2a + 2b + 2c)} * mean(q)$$

$$U_{22r} = -\frac{(-2ab - 2ac - 2bc + a^2 + b^2 + c^2)}{(2a - 2b + 2c)} * mean(q)$$

$$U_{33r} = -\frac{(-2ab - 2ac - 2bc + a^2 + b^2 + c^2)}{(2a + 2b - 2c)} * mean(q)$$

Where

$$a = \frac{\partial x_{ncorn}^*}{\partial p_{corn}} * \frac{\alpha_{rcorn}}{mean(Y_{rcorn})}$$

$$b = \frac{\partial x_{nsoy}^*}{\partial p_{soy}} * \frac{\alpha_{rsoy}}{mean(Y_{rsoy})}$$

$$c = \frac{\partial x_{nwheat}^*}{\partial p_{wheat}} * \frac{\alpha_{rwheat}}{mean(Y_{rwheat})}$$

The above formula ultimately means that there is a functional form at the mean which

relates U, α, and $\frac{\partial x_n^*}{\partial p}$ or that at the mean of the data being observed, $U_r = U\left(\alpha_r, \frac{\partial x_n^*}{\partial p}\right)$. Hence, a

conditional distribution can be formed as $U_{iir} \sim N(U_{ii}\left(\alpha_r, \frac{\partial x_n^*}{\partial p}\right), 50)$. A normal distribution was

used so that the conditional density for U is centered at this mean calculation of $U\left(\alpha_r, \frac{\partial x_n^*}{\partial p}\right)$. The

relative value of $U\left(\alpha_r, \frac{\partial x_n^*}{\partial p}\right)$ is normally a value lower than U_{ii} found nationally, between 1000-

3000. Unlike the previous chapter, which considered a larger variance on U_{ii}, the current range

was based on the uncertainty as to what the overall price response should be for the supply of the

various crops nationally. Thus if the regional slope is already given by $\left(\alpha_r, \frac{\partial x_n^*}{\partial p}\right)$ then only a

relative small range for U_{iir} needs to be considered, thus σ=50. It is assumed that the conditional

draws U_{iir} are independent between crops, so $g_U(U_r) = \prod_{i=1}^{i=3} g_{U_{ii}}(U_{iir})$.

Once the conditional draws on U are formed, then the last piece in forming this prior can be done. These are the conditional draws on h are based on U which could be defined as $\prod_{r=1}^{4} g_h(h_r|U_r)$. The conditional distribution can be found by using the functional relationship between h and U measured at the mean which is:

$$h_{ir} = \frac{\frac{mean(x_r^{obs})mean(q)}{U_{iir}} - mean(revenue_r^{obs}) + mean(cost_r^{obs}) + mean(\lambda_r^{obs})}{-mean(q)}$$

The above calculation is possible since U is diagonal, meaning there are no cross production terms that need to be considered. Using this functional form, a distribution can be formed $h_{ir} \sim N(h_{ir}(U_{iir}), 0.25)$. The relative value of $h_{ir}(U_{iir})$ is similar to national and is normally in the ranges of -5 and -15. The basic idea is that the conditional prior should be centered at this mean function and its variance is relatively small, due to small changes swinging the results of x(θ) into an infeasible range. The overall units of measure for h_i would be in dollars per acre, while Q_i would be in dollars per acre squared, and thus elements of U_{ii} would be the square root inverse of that.

Thus a prior density is formed:

$$g(\theta) = \prod_{i=1}^{3} dirch(\alpha_i) * g_{slope}\left(\frac{\partial x_n^*}{\partial p}\right) * \prod_{r=1}^{4} g_{U_r}\left(U_r|\alpha_{corn}, \alpha_{soy}, \alpha_{wheat}, \frac{\partial x_n^*}{\partial p}\right) * \prod_{r=1}^{4} g_h(h_r|U_r)$$

The likelihood function is the probability of observing each regional data point conditional on the unknown parameters. One general assumption that is needed is independence across regions conditional of the unknown parameters. It is relatively safe to assume that the unknown parameters that influence how one region allocates land will not influence how other regions should allocate land, provided we assume that farmers are price takers. Likewise the parameters influencing the shadow price of land in one region are probably not impacting the

shadow price of land in another region. Similar to the national model, the likelihood within each region should be treated as only conditional on the unknown parameters and this implies independence across crops. Overall the likelihood is:

$$f(z|\theta) = \prod_{r=1}^{4} f_{x_r}(x_r^{obs}|\theta_r) f_{\lambda_r}(\lambda_r^{obs}|\theta_r)$$

Using the previous assumptions from the national model gives $f_{x_r}(x_r^{obs}|\theta_r) =$ $\prod_{j=1}^{3} f_{x_{rj}}(x_{rj}^{obs}|\theta_r)$ where j represents the three crops. Each of the distributions $f_{x_{rj}}(x_{rj}^{obs}|\theta_r)$ and $f_{\lambda_r}(\lambda_r^{obs}|\theta_r)$ are truncated normal with mean at $x_{rj}(\theta)$, and $\lambda_r(\theta)$ respectively, truncated at 0 and with standard deviation set at x_{rj}^{obs} and λ_r^{obs} respectively allowing for a wide range to be considered.

Finally, all that is needed is to find $g(\theta|z)$ based on these distributions. The method used to find $g(\theta|z)$ is an MCMC algorithm, namely the Metropolis-Hasting algorithm. This method was applied successfully in the national model and can be used for the regional model as well. Utilizing regards to the MH algorithm for this problem, it should start with θ^0, then for j=1,2......

1. . Generate$(\theta^{j^*}) \sim J_j(|\theta^{j-1})$ where $J_j(|\theta^{j-1})$ are draws from a Dirichlet distribution vector set to α^{j-1} with that and a random draw from posterior density for the national slope and conditional draw on U is formed and based on that U a conditional draw on l is formed. Due to the fact that a Dirichlet distribution is used in the draws in the jumping distribution we have to take into account the difference in probability from going from θ^{j-1} to θ^{j^*} and from θ^{j^*} to θ^{j-1}. This changes the computation to include the jumping distribution into the formulation of r_j.

2. Compute

$$r_j = \frac{f_x(x^{obs}|\theta^j) f_\lambda(\lambda^{obs}|\theta^j) g(\theta^j) J_j(\theta^{j-1}|\theta^j)}{f_x(x^{obs}|\theta^{j-1}) f_\lambda(\lambda^{obs}|\theta^{j-1}) g(\theta^{j-1}) J_j(\theta^j|\theta^{j-1})}$$

and generate

$$W_j \sim Bernoulli(min(1, r_j))$$

3. Take

$$(\theta^j) = W_j(\theta^j) + (1 - W_j)(\theta^{j-1})$$

The important point is that this algorithm generates a point to jump to, then either accepts the point and jumps to it, or rejects it and stays at its current point. This algorithm was coded using R, how it was coded can be found in the Appendix.

2.6 Preliminary view of the Data

The data used here is total regional acreage planted and prevented acres obtained from the NASS from 2009-2013 for the crops of corn, soybeans, and wheat. The total acreage includes the prevented acreage (acres that would have been planted, but were prevented most likely due to natural disaster) to better model the farmer's decision in planting acres. The data collected from NASS were by county and then sorted into regional data. The shadow price of land was found by taking an average of expected profit weighted by crops shares. The calculation of expected profits are explained later.

Looking at the average across all regions from Table 2.1, it is evident that from 2009-2013 the Heartland had the greatest valuation for an additional unit of land, a little under $500 per acre. One clear fact is that the regions that have a heavy focus on corn and soybean tend to have a higher value for the shadow price. Northern Great Plains and Prairie Gateway have the lowest value for the shadow price and the two focus heavily on wheat.

Data surrounding cost and revenue were obtained from USDA-ERS over the same time span. The crop's cost were measured in per acre total operating costs. These costs are seed,

fertilizer, chemicals, fuel, repairs, and other variable expenses. Revenue was found by the combination of expected prices and expected yields. Expected yield is given by a linearization of the yields, $Y_t = \alpha + \beta t$. This linearization is done to mitigate unexpected disasters like droughts that could impact yields in a way that is unplanned by farmers. The data used in the linearization were actual average yields reported by the ERS from 1980-2013. Expected price is given by the approved projected price reported by USDA Risk Management Agency.

Going region by region for each crop gives an interesting look at how farming has changed in these regions in such a short time frame as shown in Table 2.2. One thing to note is the spike in commodity prices that occurs in 2011. This spike is probably due in part to the ethanol market picking up, which in turn raises corn prices. Due to this spike, there is a shift in acres devoted to corn. It is interesting to see the difference in the acreage increase from regions between 2010 to 2012. There is a 2 million increase in the Heartland, a 1 million increase in the Northern Crescent, a 2 million increase in Northern Great Plains, and a 0.2 million decrease in Prairie Gateway. This demonstrates the overall need for regional analysis, as the price response between regions are substantially different.

Looking at soybeans in Table 2.3, there is a similar effect. Again there is an obvious spike in 2011 due to the spike in commodity prices. However in this case a spike in acres is not occurring in many regions, except in the Northern Great Plains. From Table 3, it becomes obvious that the need to properly model how the competition for land is impacting the allocation of crops, we need a regional model. Many of the regions demonstrate a spike in expected profits being linked to a drop in acreage planted.

Similar difficulties can be seen when looking at wheat in Table 2.4. Again there is a jump in commodity prices in 2011. However, looking into each region, it is almost as if the acreage

allocated to wheat and expected profits are unrelated, or even implies that there is a negative relationship between expected profits and acreage for wheat. Similar to the national model, there is again a need to try and simulate the various regional models to take into account the competition for land and make sense of the various trends. With this model, however, certain regions that focus on a specific crop can be isolated and a better understanding of the overall supply function can be found.

2.7 Results of the Regional Model

For the regional crop supply model, the posterior distribution was sampled using six different starting points for θ^0, allowing the Metropolis Hasting algorithm to iterate for five hundred thousand iterations. The starting points were found for each crop using random draws from a Dirichlet distribution with the array being the portion of agricultural land in each region relative to national aggregate acreage and the scalar being 10. This procedure allows for a reasonably scattered starting point. The posterior distribution for the quadratic and linear unknown parameters do not provide too much insight. The regional slope percentage, on the other hand, does generate some interesting information.

Similar to the national model, trying to describe the posterior distribution for the regional model is difficult. This problem is enhanced by that fact that three more variables are being used in the prior and that the distribution is the joint distribution across all regions. Still the posterior distribution can be simulated and the simulations trace and plots are shown in Figures 2.1, 2.2, 2.3, and 2.4 for the Heartland. Only the posterior distribution for Heartland is shown since all other regions show similar results in their plots. Overall the distributions are all signal peaked with the expectation of a draw done for the posterior of national, which while included in the

simulation of chain and is not part of the actual posterior distribution. In many cases the plot for

U tends to have a long tail and that is due to the fact that changes in α can have huge shifts on

where U is drawn from.

The numbers in the Table 5 indicate how the price responsiveness in each region impacts

the price responsiveness nationally. For the most part, the Heartland is explaining a lot of the

national's price responsiveness. The following tables examined the distributions of the optimum

acreage based on the posterior distribution of θ^* for each of the four regions.

Table 2.6 gives the result for the Heartland based off of θ^*. The measurement is at the

median of the distribution for the simulation, and a 95% credible interval is given in parentheses.

The regional model accurately depicts the farmer's behavior for allocating land to corn and soy.

For almost every year, the median value for the simulation is within 1% of the observed acreage

for corn and within 3% for soy. Wheat is the only crop that has an observation that appears to be

an outlier; however, this crop is small with respect to the total acreage for this region. As a result

an overall accurate depiction of this crop is difficult.

There is a similar trend for the Northern Crescent region shown in Table 2.7. The

accuracy for Northern Crescent regional model is similar to the accuracy seen in the Heartland.

This is in spite of the fact that the Northern Crescent has less than one fourth the amount of

acreage in the heartland. The other two regions focus more on wheat production over corn and

soy.

It obvious from Table 2.8 for The Northern Great Plains that the results can still be fairly

accurate for these distinctly different regions. While most of the Northern Great Plains

production centers on wheat, there are a few outliers in this simulation. Again there is an

underestimation of total acres allocated to wheat. It is around 2011 that a price spike occurs and

it appears that for the regions of Heartland, Northern Crescent, and Northern Great Plains there is a jump in price for corn and soy, and less for wheat. This effect, however, is not evident in the Prairie Gateway simulation.

Table 2.9 gives the results for the Prairie Gateway. This table shows that the Prairie Gateway and Northern Great Plains are similar in that they focus on wheat. The simulated optimal acreage for Prairie Gateway is far closer to the observed acreage. There still exists an outlier for the 2010 observation of wheat, but all other simulated distributions have a median that is extremely close to the observed acreage.

The elasticity of the posterior for each region can also be seen by looking at the median and 95% credible interval (parenthetic) from a sample distribution. Table 2.10 shows the elasticity matrix for a single year (2013). It is difficult to know how reasonable these own price elasticities are for each region, as there is no literature to compare them. The values, however, mimic what a elasticity should be, given the belief of a national elasticity. The Heartland has an elasticity relatively close to the national for both corn and soy, which make sense because about 40% of these crops are produced in the Heartland. The elasticity for the supply of wheat in the Heartland is high due to the fact the acreage is relatively small. There can be similar observations made across all other regions. Northern Crescent is similar in elasticity to the Heartland which makes sense since the shares of acreage for the various crops in each region are similar. Prairie Gateway and Northern Great Plains show a different effect since they focus on wheat. Here wheat is far more inelastic than the norm, and corn and soy here are far more elastic.

These results can also be aggregated to examine how similar they are to the national results, seen in Table 2.11. The overall aggregate model seems extremely accurate; more so than the previous national model. The overall acreages will not match up, as these four regions do not

make up all national production. There still exists an outlier for the year 2011, as it assumes a greater surge in corn and soy than what is observed. Other than that, the observed aggregate acres for each crop tends to be within 1% of the median distribution of the simulation.

Looking at the aggregate elasticity across all regions in Table 2.12, there are strong similarities to the national model. These results are similar to elasticities in the national model, being (0.306, 0.314, 0.148) for corn, soy, and wheat own price elasticity. It should be noted that the elasticities for the regional model are only for 85% of all land and are not considering the elasticities of the 15% of other regions. . Examining the literature, Hendricks, Smith, and Sumner (2014) measurement of long run own price and cross price elasticities between corn and soybeans, their estimates have an own price for corn at 0.29 and soybeans at 0.26. In the short run these elasticities are 0.40 for corn and .36 for soy. Their results have short run elasticity being greater than long run elasticity, which seems like an odd result. However, the elasticities their model found could still be seen as a relative range the own price elasticity could be. The results of this chapter's model should represent the short run, overall the elasticity shown for corn and soy regional is still more inelastic than what Hendricks, Smith, and Sumner have.The results indicate that the overall aggregation of the model fits with what a national elasticity should be. This to some degree validates the elasticities posed at a regional level, as they at least aggregate to the correct elasticity.

2.8 Convergence Test

The robustness of the model can be checked through the Brooks-Gelman-Rubin (BGR) test. This check was done for 6 sample models for five hundred thousand iterations with a one hundred thousand iteration burn in. That is, the first hundred thousand iterations were thrown out

to mitigate the impact of a starting point. This test is similar to the one done in the national model, measuring R for each parameter and then as a multivariate.

The value for R should converge to 1 as n→∞.This value will allow us to determine how many iterations are suitable for the sampling to converge. There is a small complication in that the number of parameters is quite vast, so accurately measuring the multivariate value is difficult. Thus the samples parameters were broken down into their various regions and estimated separately, shown in Table 2.13.

The rule of thumb for this test is that the chains have converged if R is at 1.1 or lower. In a multivariate case this can be harder to show as some elements might not go below 1.1. Looking at the point estimates, with an expectation of one they are all under 1.1. The one that's at the board is 1.11 which given the number of variables it close enough to be considered converged since the multivariate check is also well below 1.1 among all four regions. Thus it can be assumed that the multivariate for all the parameters will satisfy the convergence criteria. These result show there is convergence among these 6 chains.

2.9 Alternative Prior

An alternative prior was considered for a robustness check. The semi random prior assumed was $\alpha_{corn} = (0.54, 0.11, 0.09, 0.13, 0.12)$, $\alpha_{soy} = (0.54, 0.08, 0.11, 0.08, 0.18)$, and $\alpha_{wheat} = (0.54, 0.08, 0.11, 0.08, 0.18)$. Convergence for this model was found using a BGR test.

The overall posterior values for α are shown in Table 2.14.The posterior density here is close to the prior considered for the model. Looking at the aggregated model for this misspecified regional model are shown in Table 2.15. This table shows that the predictions are is

close enough to the observed acreage in each year. The regional observations show similar accuracy. Overall the results are similar to the prior posed to the regional model. This result demonstrates that it is somewhat difficult to determine the correct allocation of the national slope to regional slope. Whether it is allocated via the share of total acreage to the various crops, or an alternative allocation, produces similar results. Overall this suggests that the misspecified prior could be an issue.

2.10 Forecasting

The year of 2014 is used for prediction. It is easy enough to project the expected yields and an expected cost for 2014. The expected costs are found from trending the cost data and projecting it forward one year. Similarly, expected yields are projected forward one year and expected profits per acre for each crop are then found. The only thing missing for this forecasting is the value for the total land being allocated in 2014. For the most part, total land barely changes between years, but it does in small part as the shadow price of land increases. Taking the shares of each crop allocated in each region in the previous year of 2013 an expected shadow value of land can be found. Using that expected value for the shadow land, the change in overall value is examined between 2013 and 2014, and with that shift alongside the trend for how the shadow price of land influences total land and expected value for total land in 2014 can be found. The expected price is already given by USDA Risk Management Agency.

From this the model, we predict 2014 acreage shown in Table 2.16. These results show the overall strength of our method. Using expected values for profit and total land, and being based on five years of previous data, this model accurately forecasts the following year at a regional level. The overall closeness of the forecasted simulation to what actually occurs is

extremely close. This is a more accurate prediction than national prediction done in Chapter 1

which has its median within 3.43% for corn, 6.81% for soy, and 0.62% for wheat. Where as

regionally the median perdiction is in within 0.21% for corn, 3.1% for soy, and 1.11% for wheat.

This accuracy occurs even when looking back at 2013, when the overall allocation of land is

different from 2014. The main difference between 2013 and 2014 is the expected price in 2013,

where the expected price is $5.65, $12.87, and $8.57 for corn, soy, and wheat respectively. In

2014 there's a price drop to $4.62, $11.36, and $7.02 respectively. The result demonstrates the

overall power of the model, and unless there is some fundamental shift in production that

changes farmer's behavior, it will accurately forecast how much land the farmer is willing to

allocate these three crops.

2.11 Conclusion

The work done in this chapter has shown how the previous method used to determine a

national supply model, can be used to determine a regional supply model. The general results of

the model show a lot of promise with this type of method being able to find regional elasticities,

even when information on what they should be is little or even nonexistent. The strength of this

method is how a simulation can forecast the next year using only five years of data, along with

some expected values to determine the expected profit. The accuracy of the forecast

demonstrates the usefulness of the programming model being able to accurately forecast the next

year, even when the environment and overall observed acreage compared to the previous year is

different.

The next chapter will examine the regional and national model's response to shocks. As

these are simulations for a distribution, the overall difficulty of inputting a shock is minimal. The

only real hurdle is determining a correct demand response to find the equilibrium solution. The results of the model demonstrates the usefulness of the method with the ability to form a regional model the can accurately forecast (all with limited information) in both its observations and in what the elasticity should be. This regional model is the first attempt to form a regional elasticity for these crops using US data.

2.12 References

Fernandez-Cornejo, J., & Wechsler, S. (2012). Revisiting the impact of Bt corn adoption by US farmers. Agricultural and Resource Economics Review, 41(3), 377.

Garnache, Cloé, and Pierre R. Mérel. "What can acreage allocations say about crop supply elasticities? A convex programming approach to supply response disaggregation." Journal of Agricultural Economics 66.1 (2015): 236-256.

Goodwin, B. K., & Mishra, A. K. (2006). Are "decoupled" farm program payments really decoupled? An empirical evaluation. American Journal of Agricultural Economics, 88(1), 73-89.

Goodwin, B. K., Vandeveer, M. L., & Deal, J. L. (2004). An empirical analysis of acreage effects of participation in the federal crop insurance program. American Journal of Agricultural Economics, 86(4), 1058-1077.

Henderson, J. (2008). Will Farmland Values Keep Booming?. Economic Review, 93(2), 81-104.

Hendricks, N., Smith, A., and Sumner, D. Crop Supply Dynamics and the Illusion of Partial Adjustment. Amer. J. Agr. Econ 96 (5): 1469-1491, May 2014.

Howitt, R. Positive Mathematical Programming. Amer. J. Agr. Econ. 77: 329-342, May 1995.

Jansson T. and Heckelei T. Estimating a Primal Model of Regional Crop Supply in the European Union. Journal of Agricultural Economics, Vol.62, No.1, 2011.

Nehring, R., Fernandez-Cornejo, J., & Banker, D. (2005). Off-farm labour and the structure of US agriculture: the case of corn/soybean farms. Applied Economics, 37(6), 633-649.
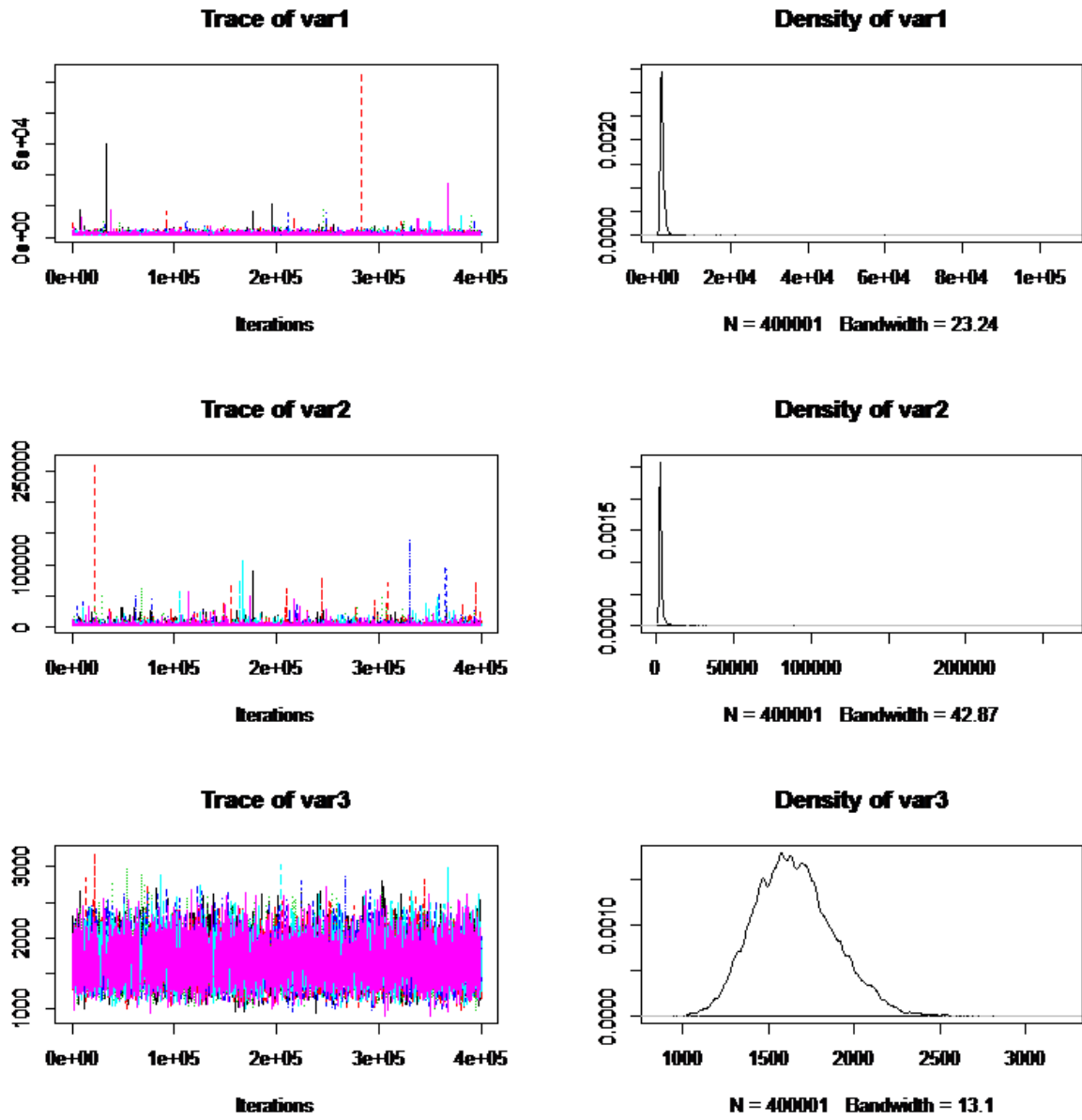
Figure 2.1 Posterior Density and Trace Plot of Regional Model for Heartland Part 1; Var1=U_{11}, Var2=U_{22}, Var3=U_{33}
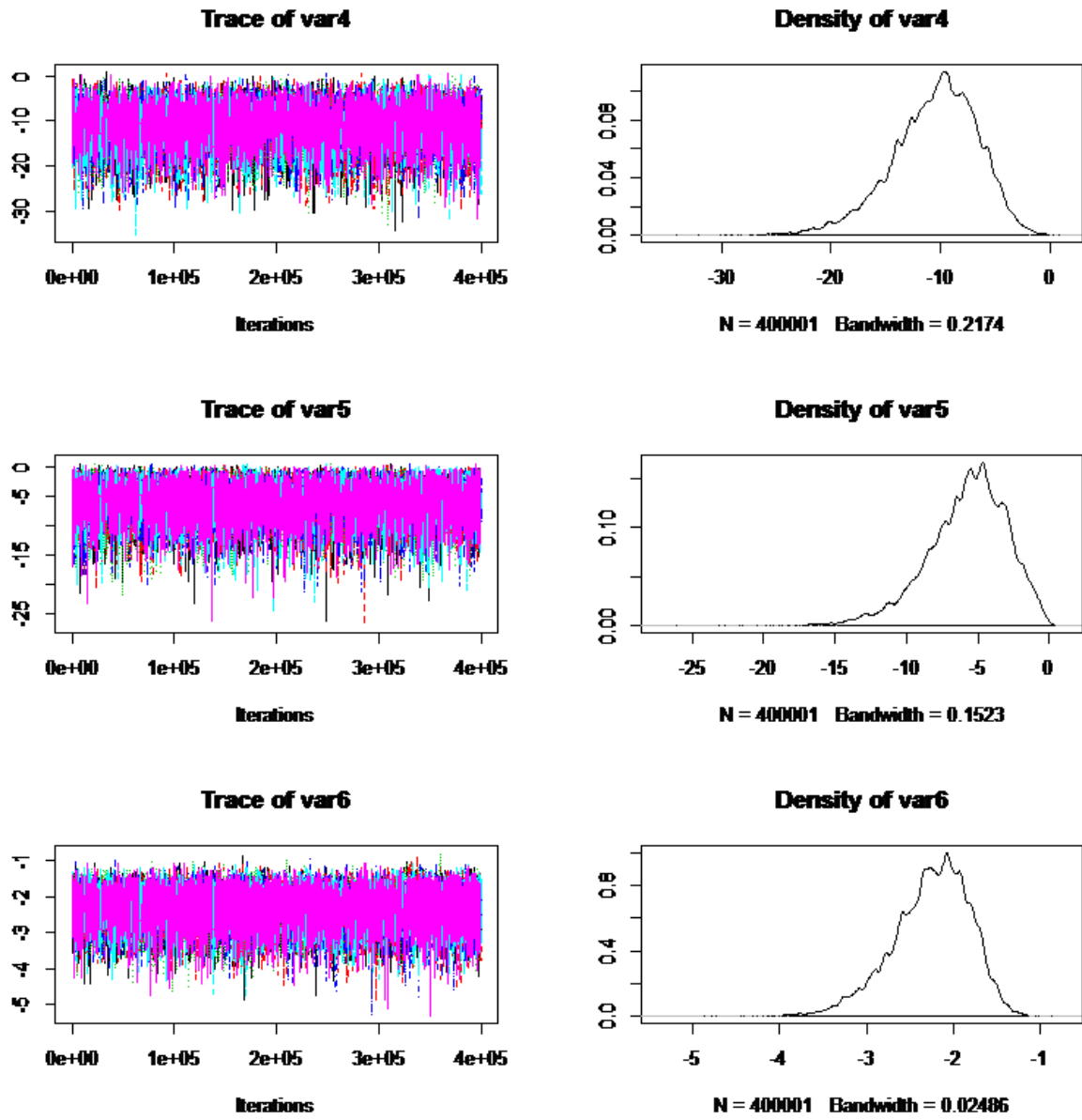
Figure 2.2 Posterior Density and Trace Plot of Regional Model for Heartland Part 2; Var4=h_1, Var5=h_2, Var6=h_3
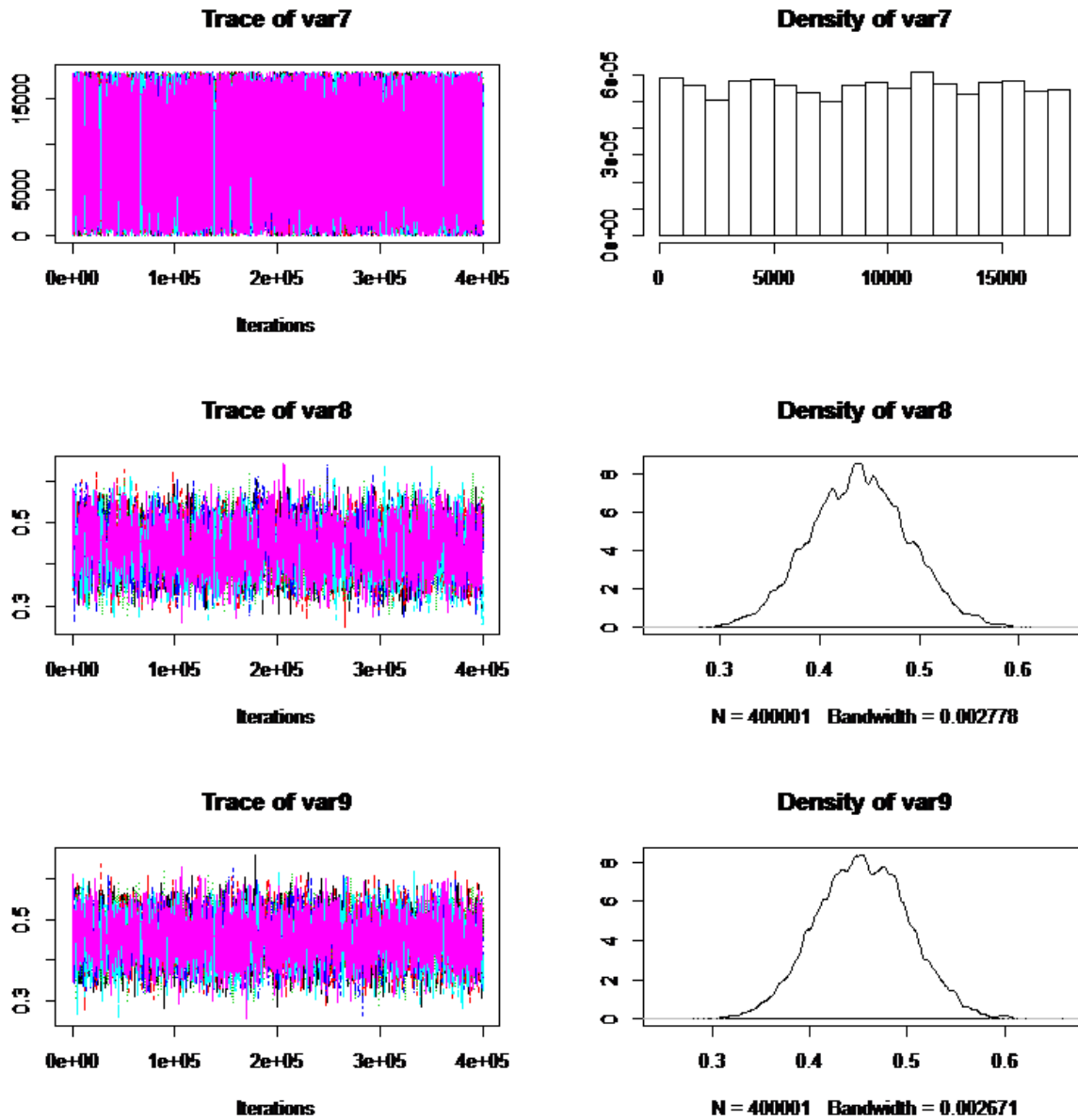
Figure 2.3 Posterior Density and Trace Plot of Regional Model for Heartland Part 3; Var7=Random draw for slope encoded in chain but not used for posterior Var8=α_{corn}, Var9=α_{soy}
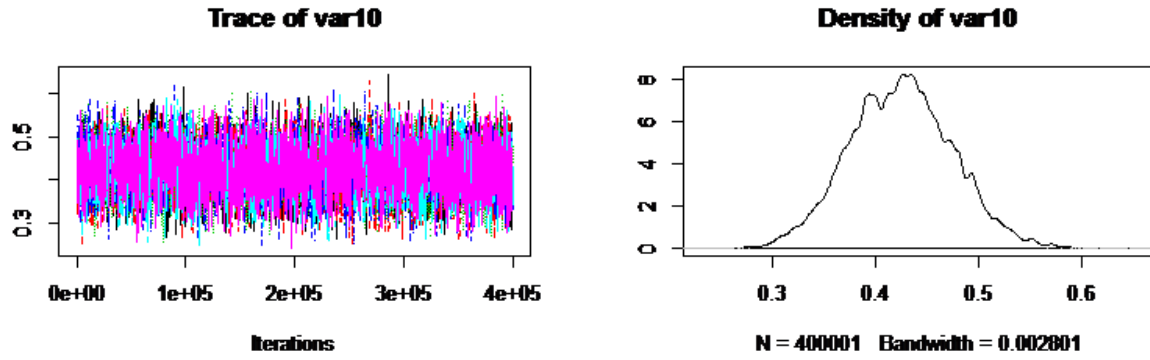
Figure 2.4 Posterior Density and Trace Plot of Regional Model for Heartland Part 4;
Var9=α_{wheat}

Table 2.1: Average Shadow Price of Land and Acreage (in Millions) across four regions Average across 5 years					
Region	Shadow Price	Corn Acres	Soy Acres	Wheat Acres	Total Acres
Heartland	$479	50.58	42.38	3.13	96.10
Northern Crescent	$367	11.14	6.42	1.43	18.99
Northern Great Plains	$235	8.32	8.55	17.44	34.31
Prairie Gateway	$266	12.22	6.29	21.80	40.31

Table 2.2: Expected Profit and Acreage (in Millions) for Corn across four Regions								
	Heartland		Northern Crescent		Northern Great Plains		Prairie Gateway	
Year	Exp. Profit	Acreage	Exp. Profit	Acreage	Exp. Profit	Acreage	Exp. Profit	Acreage
2009	$376	48.778	$254	10.530	$242	6.881	$238	11.469
2010	$380	49.275	$277	10.547	$218	7.169	$255	12.523
2011	$686	50.946	$528	11.142	$435	8.319	$481	12.638
2012	$623	52.175	$470	11.677	$388	9.196	$426	12.317
2013	$623	51.745	$472	11.809	$390	10.031	$416	12.16

Table 2.3: Expected Profit and Acreage (in Millions) for Soy across four Regions								
	Heartland		Northern Crescent		Northern Great Plains		Prairie Gateway	
Year	Exp. Profit	Acreage	Exp. Profit	Acreage	Exp. Profit	Acreage	Exp. Profit	Acreage
2009	$308	42.976	$218	6.381	$172	8.034	$160	6.009
2010	$331	43.135	$239	6.573	$184	8.319	$170	6.838
2011	$540	41.941	$409	6.371	$323	8.259	$308	6.371
2012	$464	41.611	$344	6.513	$271	8.916	$264	6.405
2013	$482	42.253	$358	6.278	$281	9.204	$275	5.809

Table 2.4: Expected Profit and Acreage (in Millions) for Wheat across four Regions								
	Heartland		Northern Crescent		Northern Great Plains		Prairie Gateway	
Year	Exp. Profit	Acreage	Exp. Profit	Acreage	Exp. Profit	Acreage	Exp. Profit	Acreage
2009	$281	3.625	$432	1.588	$221	18.315	$262	23.092
2010	$140	2.748	$196	1.290	$107	17.709	$137	22.247
2011	$204	3.146	$318	1.586	$162	19.163	$197	20.145
2012	$250	2.696	$384	1.298	$198	16.355	$236	22.114
2013	$275	3.438	$407	1.393	$215	15.658	$252	21.40

Table 2.5: Posterior Distribution of the percentage of the national slope across all regions at the Median and 95% Credible Interval

Region	Corn	Soy	Wheat
Heartland	0.44	0.454	0.425
	(0.344 , 0.537)	(0.362 , 0.548)	(0.329 , 0.525)
Northern Crescent	0.0628	0.0574	0.076
	(0.0309 , 0.1108)	(0.0280 , 0.1029)	(0.0363 , 0.1391)
Northern Great Plains	0.1466	0.1397	0.1501
	(0.0862 , 0.2179)	(0.0881 , 0.2027)	(0.0866 , 0.2326)
Prairie Gateway	0.176	0.165	0.175
	(0.112 , 0.254)	(0.106 , 0.234)	(0.107 , 0.262)

Year	Corn Predicted	Corn Obs	Soy Predicted	Soy Obs	Wheat Predicted	Wheat Obs
2009	48.92	48.78	41.55	42.98	4.84	3.63
	(45.55 , 51.91)		(38.23 , 44.98)		(3.357 , 6.880)	
2010	49.08	49.27	42.75	43.13	3.283	2.75
	(45.80 , 52.00)		(39.63 , 46.29)		(1.993 , 4.937)	
2011	51.64	50.95	42.9	41.94	1.4377	3.15
	(48.60 , 54.84)		(39.54 , 46.32)		(0.1211 , 3.0411)	
2012	51.65	52.17	42.17	41.61	2.613	2.7
	(48.57 , 54.85)		(38.70 , 45.48)		(1.381 , 4.239)	
2013	52.06	51.75	42.68	42.25	2.686	3.44
	(48.81 , 55.24)		(39.09 , 46.06)		(1.414 , 4.359)	

Table 2.6: Distribution of Predicted Crop Acreages in Millions 2009-2013 for Heartland at the Median and 95% Credible Interval

Table 2.7: Distribution of Predicted Crop Acreages in Millions 2009-2013 for Northern Crescent at the Median and 95% Credible Interval					
Year	Corn Predicted	Corn Obs	Soy Predicted	Soy Obs	Wheat Predicted	Wheat Obs
2009	10.507	10.53	6.272	6.38	1.662	1.59
	(9.916 , 11.126)		(5.728 , 6.940)		(1.411 , 2.185)	
2010	10.591	10.55	6.429	6.57	1.347	1.29
	(9.998 , 11.309)		(5.853 , 7.181)		(1.048 , 1.660)	
2011	11.34	11.14	6.49	6.37	1.2714	1.59
	(10.79 , 12.03)		(5.842 , 7.100)		(0.8399 , 1.5082)	
2012	11.59	11.68	6.49	6.51	1.44	1.3
	(10.96 , 12.25)		(5.730 , 7.074)		(1.067 , 1.724)	
2013	11.54	11.81	6.51	6.28	1.458	1.39
	(10.90 , 12.19)		(5.766 , 7.101)		(1.098 , 1.747)	

Table 2.8: Distribution of Predicted Crop Acreages in Millions 2009-2013 for Northern Great Plains at the Median and 95% Credible Interval					
Year	Corn Predicted	Corn Obs	Soy Predicted	Soy Obs	Wheat Predicted	Wheat Obs
2009	7.905	6.88	8.962	8.03	16.08	18.31
	(6.628 , 10.008)		(7.193 , 10.883)		(15.22 , 17.95)	
2010	7.884	7.17	9.491	8.32	15.59	17.71
	(6.547 , 10.074)		(7.592 , 11.510)		(14.76 , 17.20)	
2011	9.056	8.32	9.434	8.26	17.18	19.16
	(7.691 , 10.922)		(7.749 , 11.181)		(16.06 , 18.22)	
2012	8.633	9.2	9.135	8.92	16.52	16.36
	(7.323 , 10.637)		(7.434 , 10.916)		(15.66 , 17.87)	
2013	8.643	10.03	9.175	9.2	16.92	15.66
	(7.324 , 10.535)		(7.516 , 10.895)		(16.04 , 18.25)	

Table 2.9: Distribution of Predicted Crop Acreages in Millions 2009-2013 for Prairie Gateway at the Median and 95% Credible Interval						
Year	Corn Predicted	Corn Obs	Soy Predicted	Soy Obs	Wheat Predicted	Wheat Obs
2009	11.809	11.47	5.019	6.01	23.89	23.09
	(8.994 , 13.736)		(2.286 , 7.368)		(22.68 , 25.33)	
2010	12.475	12.52	5.234	6.84	24.17	22.25
	(9.425 , 14.494)		(2.315 , 7.716)		(22.44 , 25.44)	
2011	12.46	12.64	5.6	6.37	21.17	20.14
	(10.78 , 14.17)		(3.545 , 7.395)		(19.68 , 22.22)	
2012	12.79	12.32	4.974	6.4	23.28	22.11
	(10.35 , 14.68)		(2.244 , 7.145)		(21.65 , 24.50)	
2013	12.1	12.17	5.462	5.81	21.92	21.4
	(10.17 , 13.75)		(3.302 , 7.301)		(20.62 , 23.04)	

Table 2.10: Distribution of Predicted Crop Own Price Elasticity for 2013 Across all Regions at the Median and 95% Credible Interval			
Region	Corn Own Price Elasticity	Soy Own Price Elasticity	Wheat Own Price Elasticity
Heartland	0.283	0.264	1.718
	(0.179 , 0.432)	(0.165 , 0.400)	(0.861 , 3.522)
Northern Crescent	0.1853	0.2207	0.573
	(0.0818 , 0.3618)	(0.0957 , 0.4450)	(0.224 , 1.357)
Northern Great Plains	0.52	0.346	0.0876
	(0.290 , 0.886)	(0.210 , 0.578)	(0.0452 , 0.1559)
Prairie Gateway	0.469	0.739	0.0862
	(0.255 , 0.844)	(0.371 , 1.636)	(0.0449 , 0.1540)

	Corn Acres	Corn Acres Obs.	Soy Acres	Soy Acres Obs.	Wheat Acres	Wheat Acres Obs.
Table 2.11: Distribution of Predicted Crop Acreages in Millions 2009-2013 for the Aggregation across the four Regions at the Median and 95% Credible Interval						
2009	79.16	77.66	61.86	63.40	46.62	46.62
	(74.63 , 83.24)		(57.13 , 66.44)		(44.31 , 49.94)	
2010	80.07	79.51	63.92	64.86	44.42	43.99
	(75.46 , 84.19)		(59.24 , 68.69)		(42.13 , 47.02)	
2011	84.59	83.04	64.41	62.94	41.01	44.04
	(80.84 , 88.77)		(60.10 , 68.64)		(38.61 , 43.24)	
2012	84.69	85.36	62.74	63.44	43.87	42.46
	(80.63 , 88.93)		(57.94 , 67.02)		(41.66 , 46.22)	
2013	84.38	85.75	63.81	63.54	43.01	41.89
	(80.38 , 88.43)		(59.26 , 67.99)		(40.95 , 45.34)	

Table 2.12: Distribution of Predicted Crop Own Price Weighted Average Across the Four Regions at the Elasticity for 2013 Median and 95% Credible Interval			
	Corn	Soy	Wheat
Corn	0.306	-0.1503	-0.0618
	(0.207 , 0.433)	(-0.2357 , -0.0902)	(-0.1388 , -0.0177)
Soy	-0.336	0.314	-0.1691
	(-0.561 , -0.198)	(0.210 , 0.452)	(-0.3844 , -0.0565)
Wheat	-0.902	-1.037	0.1475
	(-2.466 , -0.185)	(-2.412 , -0.376)	(0.0997 , 0.2050)

Table 2.13: measurement for R over all theta										
Region	U_{11}	U_{22}	U_{33}	h_1	h_2	h_3	α_{corn}	α_{soy}	α_{wheat}	Multi
Heartland	1.00	1.11	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01
Northern Crescent	1.02	1.06	1.01	1.00	1.00	1.01	1.00	1.00	1.00	1.01
Northern Great Plains	1.03	1.03	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Prairie Gateway	1.02	1.02	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 2.14: Posterior Distribution of the percentage of the national slope across all regions at the Median and 95% Credible Interval (Misspecified)			
Region	Corn	Soy	Wheat
Heartland	0.596	0.581	0.553
	(0.497 , 0.686)	(0.488 , 0.669)	(0.453 , 0.653)
Northern Crescent	0.0883	0.0705	0.0779
	(0.0488 , 0.1409)	(0.0361 , 0.1176)	(0.0347 , 0.1430)
Northern Great Plains	0.0834	0.0937	0.1105
	(0.0382 , 0.1475)	(0.0538 , 0.1390)	(0.0554 , 0.1872)
Prairie Gateway	0.1097	0.0743	0.0747
	(0.0643 , 0.1681)	(0.0342 , 0.1242)	(0.0341 , 0.1350)

Table 2.15: Distribution of Predicted Crop Acreages in Millions 2009-2013 for the Aggregation across the four Regions at the Median and 95% Credible Interval (Misspecified)						
	Corn Acres	Corn Acres Obs.	Soy Acres	Soy Acres Obs.	Wheat Acres	Wheat Acres Obs.
2009	77.94	77.66	62.73	63.40	46.85	46.62
	(72.72 , 82.97)		(57.12 , 68.11)		(44.52 , 50.57)	
2010	78.58	79.51	64.95	64.86	44.74	43.99
	(73.12 , 83.75)		(59.39 , 70.41)		(42.80 , 47.75)	
2011	84.13	83.04	64.71	62.94	41.11	44.04
	(79.58 , 88.99)		(59.72 , 69.48)		(39.34 , 43.24)	
2012	83.65	85.36	63.41	63.44	44.08	42.46
	(78.56 , 88.87)		(58.08 , 68.57)		(42.32 , 46.53)	
2013	83.74	85.75	64.18	63.54	43.17	41.89
	(78.88 , 88.61)		(59.02 , 69.19)		(41.53 , 45.56)	

Region	Corn Prediction	Corn Obs	Soy Prediction	Soy Obs	Wheat Prediction	Wheat Obs
HL	49.92 (46.39 , 52.90)	49.65	43.60 (40.30 , 47.18)	44.17	3.25 (1.928, 4.95)	2.84
NC	10.88 (10.23 , 11.42)	10.79	6.53 (5.99, 7.21)	6.74	1.48 (1.20 , 1.81)	1.22
NGP	7.97 (6.61 , 9.89)	8.26	9.70 (7.94 , 11.63)	10.85	16.33 (15.47 , 17.78)	17.06
PG	11.62 (9.54 , 13.17)	11.51	6.24 (4.29 , 8.19)	6.49	21.64 (20.24 , 22.69)	22.1
Agg	80.39 (76.02, 84.20)	80.22	66.13 (61.88, 70.74)	68.25	42.73 (40.66, 45.16)	43.21

Table 2.16: Distribution of Prediction for Acreage (in Millions) for the Regions for out of Sample Year 2014 at the Median and 95% Credible Interval

CHAPTER III: IMPACTS ON NATIONAL AND REGIONAL CROP SUPPLY MODELS

3.1 Abstract

In the previous chapters we have been able to accurately form a national and regional model for the allocation of land for corn, soy, and wheat for the US using only five years of data. This type of model allowed for quick and simple forecasting for future years that only requires expected prices and costs. This chapter will exam these models in a new setting by examining the effect of shocks on the system. Two types of shocks will be considered, one a simple external price shock on a crop. The second impact considered is a shock in yields for a given region simulating what could potentially occur in a drought, for example. These types of shocks will be examined at the equilibrium using a simple model to represent the demand for the crops.

3.2 Introduction

In general, models focusing on modeling a supply of commodities are trying ultimately to find determinants for the supply of crops or price of the crop, forecast future supply, or determine how various shocks could impact the model. The previous chapters created a model that could successfully forecast future supply of corn, soy, and wheat at the national and regional level using only five years of data. These models employing Positive Mathematical Programming and Bayesian Analysis do not in any way identify what determinants could impact the crop supply other than a somewhat simple view of average profit per acre per crop. The previous work does create calibrating parameters such that even if we do not have full understanding as to what parameters are impacting this model, one can simulate a reasonable solution. With proper beliefs as to the form of the second order condition, the model can take

into account price effects and forecast future land allocations. Ultimately the strength of these models is that they are able to forecast in an accurate way with a limited data set.

The next step is to examine how this model reacts to shocks. In this analysis the modeling of shocks will be compared to other literature to determine how reliable this type of analysis is. We find results similar to those from other relevant literature. The unique type of modeling demonstrated here makes it relatively easy to see how the impact of a price decrease in corn can impact other crops, such as wheat and soy. This type of experiment appears to be a quick analysis of price shocks, needing little information, and is able to forecast and simulate a shock with respect to that forecast quickly and precisely. More notable, this type of modeling can also show the impact of shocks not just nationally, but regionally. While the national and regional models may differ slightly, that is mainly due to the use of different data sets for the two models.

This chapter examines the effects on price and supply of corn, soy, and wheat at an equilibrium point. Forming a supply curve from the previous chapter and using a simple constant elasticity for demand for the three crops to get an accurate view of how a price shock can affect the model will do this. This model projects forward using 2014 forecasts to determine the impacts of future shocks in this model. The model will examine what occurs with a price decrease and increase as well as a yield shock in corn in 2014. These results are then compared with other relevant literature that examines shocks.

3.3 Literature Review of Impact Models

There have been a variety of impact models examining both national and regional implications. These models were constructed for a variety of reasons. Many of the more recent studies stem from the desire to understand the impact of ethanol on the price of corn. An article

by Naylor, et al (2007) lists the reasons to be concerned about the rise in biofuels with food insecurity worldwide. They illustrate the fact that as of 2007, the US makes up 40% of world production of corn and around 60% of global trade. They go on to show the pressure put on corn impacts all crops, lessening the acres allocated to crops like soy and wheat to produce more corn. These increases in prices for the various crops used in biofuels alongside other crops mentioned would then lead to greater food insecurity in poor nations. They concluded these trends would occur as long as the price for crude oil remains high, increasing the demand for ethanol. Larger impacts would be due to the constraint put on agricultural land leading to either an expansion of land use in agriculture or an increase of intensities on inputs to increase yields. However, the impact model considered for this chapter will be for a drop in corn prices. This is due to the fact that the year that this model considers, 2014, is the height of biofuel demands and there is evidence of dropping crude oil prices. This brings up an interesting question on how such a decrease will impact production.

The use of a PMP model studying impacts is nonexistent. There has been a study on forecasting by Kanellopous et al. (2010) with the Farm System SIMulator. The method used in their paper is different from what is described here, requiring some modification for use with time series. However, they found that the use of a PMP variant created forecasts close to actual farmer's decisions compared to alternative methods. The overall strength of the forecasting done in the previous chapters support this claim, but it is still unknown how well PMP works in an impact model.

Some literature does exist using regional information to help determine various impacts. The study that best highlights this is Fernandez-Cornejo and Wechsler (2012). They looked at determining the impact of using Bt corn (genetically engineered crop with Bacillus thuringiensis)

on the profits and the use of insecticides on farms. The overall impact they found was that Bt seed would increase profit, yield, and seed demand. Their use of regional information is somewhat limited in that only the Heartland was used as a dummy variable. Their finding was that farmers in the Heartland are more likely to adopt Bt seed, but not to a significant degree. An alternative method similar to the regional model posed in the previous chapter might make it possible to better understand how the adoption Bt corn effects the various regions.

Drought impact studies have also be on the rise, due in part to fears surrounding climate change. Wilhite, Svoboda, and Hayes (2007) highlight the fact that there is an absence of a database for studying the impacts of droughts. Wilhite et, al. established a database in 2005 and highlight concerns over the future frequency of droughts. Their study examines the increase in areas suffering droughts from 2000 to 2005 and the increase in crop indemnities from crop insurance due to the droughts. There is a startling amount of costs nationally surrounding a drought with few existing studies on the issue. Thus there is a need for impact studies on how a drought can effect crop production. We will show how our model can used to simulate these results.

Another study done by Tokgoz, et. al (2008) examined how the price for corn could shift in 2016, by increasing oil prices, or if a drought similar to 1998 occurred. The study found that a $10 increase in crude oil prices would result in a 20% increase in corn prices. Alternatively, they considered a scenario where a drought of similar magnitude to the 1988 drought occurs and found that corn prices increase by 44%, soybean prices increased by 22%, and wheat prices increases by 15%. While the overall impact of crude oil prices would not be comparable to this chapter's model, as it will not be relating crude oil price to the demand for corn, the drought model this study puts forth could be comparable to a certain degree.

There have been many articles regarding the impact of various policies and disasters, but few try to include any regional implications. The national and regional model examined in previous chapters makes it possible to do both. First this chapter needs to restate key findings of the previous chapters and then define a demand model.

3.4 Crop Supply Model

In the previous chapters we were successfully able to find a posterior distribution both at a national and regional level for a pair of unknown parameters (Q, h) that are used in calibrating a nonlinear model that determines the optimal allocation of land across the crops of corn, soy, and wheat. The nonlinear model is listed below:

$$\max_{x_t} x_t [p_t Y_t - cost_t] - q_t x_t' \left[h + \frac{1}{2} l_t Q x_t \right]$$

$$\text{such that } R x_t = v_t$$

(1)

This type of modeling then solves for its first order conditions, giving $x_t(Y_t, p_t, cost_t, q_t, l_t, v_t)$ which is

$$x_t^* = E_t^{-1}[p_t Y_t - cost_t - q_t h]$$

$$- E_t^{-1} R_t ([R_t E_t^{-1} R_t']^{-1} [R_t E_t^{-1} (p_t Y_t - cost_t - q_t h) - v_t])$$

(2)

Where $E_t = q_t l_t Q$. The implications is that by having a posterior density of Q and h, all that is needed to determine the acres the farmers will allocate for each crop x^* is some understanding of expected yields per acres for the crops Y, expected prices for the crops P, expected costs per acre for the crops cost, a price index variable to take account for inflation q, and the total available land for the farmers to use v. Note that the land index variable l the total

land index is to be used to take into account any shifts in the total available land between years. However this index is ultimately determined by v and is normally close to 1.

This is what makes the overall use of the model so appealing. By having the posterior distribution on Q and h, it is easy to find x^* for a future year or under varying circumstances. This method has already been used in forecasting fairly accurate results for 2014, which was substantially different in terms of expected prices for 2013. That is because this method is ultimately trying to find a functional form for production that fits the various restrictions we impose on price responsiveness. By finding this functional form, as long as production does not structurally change in a way impacting Q and h, then it should be accurate in forming x^* under any condition involving yields, prices, cost, or total acreage. However, in order to create a proper impact model, there needs to be some representation of the other side of the problem, consumer demand.

3.5 Crop Demand Model

The focus of this section is to develop a suitable model for demand for corn, soy, and wheat in order to determine how the equilibrium price and quantity shifts for the three crops. This is by no means a small feat, as these products are consumed in a variety of ways, be it raw consumption, ethanol, or for feed. There are a variety of ways a demand model could be formed, however, as the focus of these works are to demonstrate the usefulness of the crop supply model, a simple demand model is considered. A more complex demand model could be formed as was done by Arfini et al. (2008) using a PMP method, but this is beyond focus of this work. The overall model considered here for demand is:

$$QD = Ap^{\varepsilon}$$

Where QD is the quantity demand for the crop, A is some constant term, and ε is the

demand elasticity for the crop. The model is an example of a function with a constant elasticity ε.

The important aspect to this model is that it represents how the consumer reacts to the change in

quantity supplied or price of the good. As long as the elasticity chosen is close to reality, the

demand model should be accurate in showing the impact of small changes.

Looking at the literature, there is a rather broad picture of what the elasticity of demand

should be. However, finding some general elasticity that encompasses the demand for the crop is

difficult due to the fact the goods are used domestically as feed, as ethanol, and as an export.

Taking that into account, and the variety of elasticities reported on these crops, it was decided

that the demand of elasticity corn, soy, and wheat, should be around (-0.91,-0.38,-0.75). The

relative demand elasticities have not been extensively researched. Gardner (2007) states that the

elasticity of demand for feed and export of corn should be at -0.91 according to USDA. For soy,

Piggott, Nicholas, and Wohlgenant (2002) calculate a value for the elasticity demand for US soy

of -0.38. Barnes and Shields (1998) of the ERS disaggregate wheat for better estimates of

demand elasticity and find that the elasticity of demand of winter wheat (over 75% of wheat

production) of -0.75 in 1995. Winter wheat is used as all wheat in this model, because there is no

way to disaggregate wheat in the present data set. Since the prices are similar between wheat and

winter wheat, demand for winter wheat is used as it is the majority of wheat produced in the US.

While this give us an idea of what the elasticity of demand should be, the relative range could

have these elasticities go as low as -0.2 or as high as -1.

Using these elasticities, a demand model can be formed. This demand model, in

conjunction with the supply model, can give an idea on how shifts to the supply curve could

impact the overall equilibrium point. The method used in calculating the shift in equilibrium

price and quantity for a prediction referred to as Yx¹ and p¹ is as follows:

1. A shock on the expected profit for the various crops occurs

2. This shock is then used with the simulated (Q, h) and finds x^{j+1}

3. The percent changed between Yx^{j+1} and Yx^j is calculated, and then is used in the demand

model to find the percent change in prices, which gives us p^{j+1}

4. This new price is then used to calculate a new shock to be used in step (1)

This method iterates until prices are no longer changing and have converged. In general

25 iterations of this method is all that is needed for prices to converge and an equilibrium price

and quantity is found for this shift in either supply or demand. This method is done for all

samples simulating (Q, h) and so we are ultimately finding this shift in equilibrium price and

quantity for all variations in (Q, h) that we have simulated. Thus this model is simulating the

change in the equilibrium price and quantity, giving us a distribution for where this new

equilibrium could be. Because of this, there can be problems with outliers not converging to an

equilibrium point. This problem is mainly due this method causing non-convergence if the

elasticity of supply is sufficiently inelastic when compared to demand elasticity. While the

outliers cause a problem in some of the samples, causing a wider range for this simulation, the

median is where most of the distribution is congregated and is a fair representation of this new

equilibrium point. This method described here is the same for both national and regional models.

However, in the regional model, when calculating the percent change from Yx^{j+1} to Yx^j we

look at the aggregated values. Thus, for the acreage found regionally, x_r, we look at the percent

between $\sum_{r=1}^{4} Yx^{j+1}$ and $\sum_{r=1}^{4} Yx^j$. A better demand model could give a more accurate

examination as to how the market changes due to external changes. But for this example, we can

see if a simpler modeling of demand is realistic, compared to other literature in this area. This algorithm was coded for both national and regional using R, how it was coded for regional can be found in the Appendix.

3.6 Results for the Impact Model

The year that will be examined for this impact modeling will be the forecasted year of 2014. As most of the modeling of this nature is used for future years, and as 2014 has already be forecasted in the previous chapter both nationally and regionally, this year will be used for the impact model. To reiterate how the forecasting was done both nationally and regionally, first expected yield and expected cost were formed. The expected costs are found from linearizing the cost data per acre both nationally and regionally for each crop from 2009 to 2013, then projecting it forward one year. Similarly, expected yields are linearized from 1980-2013 and then are projected forward one year and expected profits per acre for each crop are then calculated. The last part is determining total land acreage for 2014 both regionally and nationally. For the most part, total land barely changes between years, but it does in small part as the shadow price of land increases nationally. A dollar increase in the value of the shadow price of land increases total acres by 46510. Regionally there is an increase of (4,569, 4,378, 6,773, and -8,248) for Heartland, Northern Crescent, Northern Great Plains, and Prairie Gateway respectively (note due to a negative relationship for Prairie gateway, zero was assumed). These numbers are mostly negligible compared to the millions of planted acres. These values are used against the change in the shadow price to determine how much total acreage should increase from 2013. Using this, forecasting for 2014 acreage was done.

The results nationally can be seen in Table 3.1. The overall forecasting is within the credible intervals for two of the three crops. There is a clear difficulty in trying to predict soy and so a more accurate model is attempted by using more data at a regional level. Attempting a regional estimation on the four regions of Heartland, Northern Crescent, Northern Great Plains, and Prairie Gateway yields Table 3.2. This is by far a more accurate estimation than the national model. As explained in the previous chapter there is a gain in accuracy as we move from the national to the regional model. This accuracy occurs even with a real price drop occurring in 2013 to $4.62, $11.36, and $7.02 for corn, soy, and wheat respectively. With these predictions for 2014 we can then see how the model reacts to changes we impose on it by applying a shock on expected profits.

Suppose we are an analyst in 2013 and wanted to determine the impact of a policy that would cut the demand of ethanol in 2014. Additionally we had knowledge that the initial price shock from the policy could decrease the price of corn by 20%, or from $4.62 to $3.70. The question is then, how would farmers both nationally and regionally react to such a price drop, and what would be the price and crop allocation when it returns to an equilibrium? By using the supply function both nationally and regionally and a national demand function, we can apply this shock and iterate between supply and demand, until an equilibrium point is found.

Nationally this result would be shown in Table 3.3. This basic result, while having a somewhat wide range for the credible interval, is rather dense around the median point. This fits with what would be assumed for an equilibrium price for corn, somewhere between the original shock and the initial price shock. In addition, for the most part, wheat and soy have a decrease in price due to the fact that farmers are moving away from corn production and into the production of soy and wheat. Looking at the median value, for the most part, these numbers agree with what

one would assume with a price shock on corn. The overall acreage goes down for corn, with a slight increase in the production of soy and wheat. Looking at the regional model, a more precise picture of the aggregate change can be formed and shown in Table 3.4. The overall ranges in this table in the aggregate is much more precise with a lower overall range. The percent change at the median for the aggregated value is approximately the same as the change in the national model.

This demonstrates that the two approaches, for the most part, are consistent in their predicted changes. Regionally, it can be a bit ambiguous as to how the overall acreage changes in each region. In all cases it is obvious that the region will decrease land allocated to corn. The Heartland shifts its production into soy and wheat, which shows a significant increase. This increase in production for soy and wheat seen in the Heartland might be making the overall change in soy and wheat in the other regions more ambiguous. There appears to be a solid increase in acres allocated to soybeans. However, wheat is somewhat more ambiguous, as shifts into soybean might be more favorable for the Northern Crescent and Northern Great Plains. The equilibrium price found with this regional model is close to the overall equilibrium price found nationally. The numbers overall seem to line up between the two models. Thus these models not only show an overall fall in price in corn from this policy, but also the fall in price for the other commodities as well. The regional model appears to show a more precise model for the impact of the policy.

In addition we will also consider what would occur if the reverse happened, and a policy increases the demand for ethanol, which causes a 20% price spike for 2014, or from $4.62 to $5.54. The national model highlighting this change is shown in Table 3.5. The overall results underline certain problems with an impact model, which need to be considered with this approach. While this method is quick and simple, it can have a convergence problem with

outliers. The entire sample is being iterated between itself and the demand equation to find the point which is at equilibrium. For a lot of the points it can converge to an equilibrium, but for some outliers that are either more elastic or more inelastic there are convergence problems, and the results go to $\pm\infty$. This would mean the overall credible intervals are larger than what they should be.

Even with the larger credible intervals, there is still a lot that can be learned looking at the median, where a great deal of the distribution is. The median values of the prices for the various crops are higher than what they are initially and overall there is a reverse than what occurs with a price decrease. Corn acres increase nationally and acres allocated to soy and wheat decrease.

Looking at the impact at a regional level, this model eliminates some of the outliers present in the national model as seen in Table 3.6: The aggregated portion of the model is more certain in terms of the change in acres and the price change, however it is matching the national model at the median value. The results at a regional level are bit more uncertain. Heartland is showing a definite increase in acres devoted to corn and a decrease in soybean and wheat acreage. The Northern Crescent is showing a definite increase in corn with some uncertainty as to how acres will change in soy and wheat. Similarly, it is difficult to say the overall acreage change for the Northern Great Plains and Prairie Gateway, as both are showing a substantial decrease in wheat and increase in corn. A comparison to these results are given by Fabiosa et. al. (2010) which used the international FAPRI Model to judge the impact of land allocation effects from a global ethanol surge. Looking at their model, if we were to assume a 20% increase in ethanol demand, that would mean that for the US, corn area goes up 2.86% , soybeans goes down 1.96%, and wheat goes down 0.46%, while prices go up by 5.76%, 1.8%, and 2.14% for

corn, soy and wheat respectively. The numbers from Table 3.6 are similar in terms of the relative

magnitudes of how it impacts the crops. While one would not expect these models to exactly

match up, there are similarities. Depending on what you would assume for the demand model, it

could be brought closer to what is posed in their model. Given the demand elasticities used in our

present model are from different time periods than the data used to model supply, it reinforces

the need to extend the present research to a more formal approach to modeling the demand side.

A model of constant elasticity for demand, although simple to implement, may not be realistic.

What is also important to consider is that the results shown in Table 3.6 were done with a very

limited data set.

There are other types of regional models which could be developed. The most obvious

one is how would an event in a region impact other regions as well as a nationally. Suppose that

instead of looking into a policy impact that influenced price, we wanted to look at a regional

impact for 2014. Suppose that there are fears centering over a drought in the Heartland, or some

sort of regional disaster that would impact overall yields. Work done by Yu and Bacbcock

(2010) relates a drought index to a loss in yields. From their study, Iowa is known to have to

have encountered droughts of an index of 2. Suppose that there is a drought of that magnitude in

2014 for the Heartland, which according to their study, would mean a loss of yields of about

20% across soy and corn. Let us extend this loss also to wheat (this number would be difficult to

measure as the Heartland does not produce as much wheat, so 20% is just assumed). It should be

noted that for this type of modeling, we would also assume that the farmers had some knowledge

of this event occurring before it happens, and are making decisions before it impacts the crops.

This is not realistic, however, if we assumed the farmers had no knowledge, their behavior would

not change and this impact would all depend on the model for demand. We could also view this

as some unavoidable drop in yield, there is knowledge of this, and the expected prices still need
to be updated to reflect it.

The Table 3.7 shows an interesting but also expected outcome. The overall acres being
allocated for the Heartland barely changes due to the fact that all crops are suffering a loss in
expected profits, with the least impacted being soy. Due to the Heartland producing less, all other
regions allocate more land to corn and soy, this nets to an overall decrease in the land allocated
for wheat. The variability seems a bit high, and might be due to overall uncertainty as to which
regions will be shifting production and by how much. Recalling the drought scenario, Tokgoz, et
al. (2008) posed, there was a 44%, 22%, and 15% increase in corn, soy, and wheat prices
respectively. This model with a moderate drought in the Heartland shows a 13%, 20%, and 3%
increase in corn, soy, and wheat prices respectively. The fact that we are making the unrealistic
assumption that the farmers could react to the drought and that it is all localized in the Heartland
might be why the percent change in prices are not as high as one would predict. This in part
because of the reaction of other regions to increase produce in corn and soy, which would lessen
the overall loss in production. This model is in some ways illustrating the value of the
information of the yield shock. While the assumptions to the yield shock model are unrealistic,
these predictions were easy enough to form. As long as the impact of a policy or event can be
related to the expected profits of the crops, an impact model can be simulated.

3.7 Conclusion

The work done in this chapter shows the relative usefulness of the regional supply model
in not only accurately forming a prediction, but it can also be used in impact simulation. The
method does calibrate a fairly accurate supply curve for the allocation of crops, and with that all

that is needed is the linking of information with respect to a crop's expected profit to determine how production might change. Our model also gives some insight into the relative reaction between regional and national impacts.

This work could be expanded upon in a variety of ways. One could create a more complex model looking at a variety of variables not considered which would influence the production function. There could also be changes to the demand function used here. One option could be the route taken by Arfini et al. (2008) and attempt to use the PMP method to find a suitable demand function. Thus far the results for this methodology has been interesting, given the limited data available, and future research can expand on this type of modeling.

3.8 Reference

Arfini, F., Donati, M., Grossi L. and Pariss, Q. Revenue and Cost Functions in PMP: a Methodological Integration for a Territorial Analysis of CAP. Paper presentation at the 107th EAAE Seminar "Modelling of Agricultural and Rural Development Policies". Sevilla, Spain, January 29th -February 1st, 2008.

Fabiosa, Jacinto F., et al. "Land allocation effects of the global ethanol surge: Predictions from the international FAPRI model." Land Economics 86.4 (2010): 687-706.

Howitt, R. Positive Mathematical Programming. Amer. J. Agr. Econ. 77: 329-342, May 1995.

Gardner, Bruce. "Fuel ethanol subsidies and farm price support." Journal of Agricultural & Food Industrial Organization 5.2 (2007).

Jansson T. and Heckelei T. Estimating a Primal Model of Regional Crop Supply in the European Union. Journal of Agricultural Economics, Vol.62, No.1, 2011.

Kasnellopoulos, A. Berentsen, P., Heckelei, T., Van Ittersum, M. and Oude Lansink. Assessing the Forecasting Performance of a Generic Bio-Economic Farm Model Calibrated with Two Different PMP Variants. Journal Of Agricultural Economics 52(7):333-340. 2010.

Barnes, James N., and Dennis A. Shields. "The growth in US wheat food demand." *Wheat Yearbook* (1998): 21-29.

Naylor, R. L., Liska, A. J., Burke, M. B., Falcon, W. P., Gaskell, J. C., Rozelle, S. D., & Cassman, K. G. (2007). The ripple effect: biofuels, food security, and the environment. Environment: Science and Policy for Sustainable Development, 49(9), 30-43.

Piggott, Nicholas E., and Michael K. Wohlgenant. "Price elasticities, joint products, and international trade." Australian Journal of Agricultural and Resource Economics 46.4 (2002): 487-500.

Tokgoz, S., Elobeid, A., Fabiosa, J., Hayes, D. J., Babcock, B. A., Yu, T. H. E., ... & Hart, C. E. (2008). Bottlenecks, drought, and oil price spikes: Impact on US ethanol and agricultural sectors. Applied Economic Perspectives and Policy, 30(4), 604-622.

Vidal, I., Iglesias, P., and Galea, M. Influential Observations in the Functional Measurement Errror Model. Journal of Applied Statistics. Vol. 34, No.10, Dec 2007.

Wilhite, D. A., Svoboda, M. D., & Hayes, M. J. (2007). Understanding the complex impacts of drought: a key to enhancing drought mitigation and preparedness. Water resources management, 21(5), 763-774.

Yu, T., & Babcock, B. A. (2010). Are US corn and soybeans becoming more drought tolerant?. American Journal of Agricultural Economics, 92(5), 1310-1323.

Table 3.1: Prediction for Crop Acreagefor 2014 (in Millions), Median & 95% Cred. Interval		
2014	Predicted	Observed
Corn	89.29 (85.44,93.29)	92.46
Soy	79.22 (75.34,83.09)	85.01
Wheat	57.84 (54.81,60.88)	58.20

Table 3.2: Distribution of Prediction for Acreage (in Millions) for the Regions for out of Sample Year 2014 at the Median and 95% Credible Interval

Region	Corn Pred	Corn Obs	Soy Pred	Soy Obs	Wheat Pred	Wheat Obs
HL	49.92 (46.39 , 52.90)	49.65	43.60 (40.30 , 47.18)	44.17	3.25 (1.928, 4.95)	2.84
NC	10.88 (10.23 , 11.42)	10.79	6.53 (5.99, 7.21)	6.74	1.48 (1.20 , 1.81)	1.22
NGP	7.97 (6.61 , 9.89)	8.26	9.70 (7.94 , 11.63)	10.85	16.33 (15.47 , 17.78)	17.06
PG	11.62 (9.54 , 13.17)	11.51	6.24 (4.29 , 8.19)	6.49	21.64 (20.24 , 22.69)	22.1
Agg	80.39 (76.02, 84.20)	80.22	66.13 (61.88, 70.74)	68.25	42.73 (40.66, 45.16)	43.21

Table 3.3: National Model Change in Prices and Acres from 20% decrease in Corn Median & 95% Cred. Interval			
Nationally	Corn	Soy	Wheat
Price	$3.84 ($3.52 , $4.10)	$10.52 ($7.28 , $12.24)	$6.75 ($6.20 , $7.52)
Change in Acres in %	-4.05% (-10.48% , -0.48%)	0.96% (-5.29% , 3.53%)	2.75% (-5.34% , 8.22%)

Table 3.4: Regional Model Change in Prices and Acres from 20% decrease in Corn Median & 95% Cred. Interval			
Change in Acres %	Corn	Soy	Wheat
Heartland	-3.44%	1.39%	22.10%
	(-4.44% , -0.93%)	(-3.82% , 2.00%)	(10.1% , 55.7%)
Northern Crescent	-4.34%	0.18%	3.34%
	(-5.79% , -2.40%)	(-3.37% , 3.76%)	(-2.20% , 14.22%)
Northern Great Plains	-5.88%	2.39%	0.05%
	(-8.95% , -2.90%)	(-3.38% , 4.67%)	(-0.95% , 2.14%)
Prairie Gateway	-4.78%	5.11%	1.16%
	(-8.53% , -1.59%)	(-7.24% , 10.04%)	(0.36% , 3.09%)
Aggregated	-4.06%	1.87%	2.39%
	(-4.87% , -1.47%)	(-4.10% , 2.27%)	(1.21% , 6.60%)
Price	$3.84	$10.49	$6.73
	($3.53 , $3.86)	($6.95 , $10.90)	($6.07 , $6.87)

Table 3.5: National Model Change in Prices and Acres from 20% increase in Corn Median & 95% Cred. Interval			
Nationally	Corn	Soy	Wheat
Price	$5.18	$10.74	$7.02
	($3.98 , $5.78)	($2.94 , $13.01)	($5.39 , $7.94)
Change in Acres in %	4.41%	-5.56%	-0.99%
	(-4.19% , 15.43%)	(-22.55% , 3.85%)	(-9.69% , 16.42%)

Table 3.6: Regional Model Change in Prices and Acres from 20% increase in Corn Median & 95% Cred. Interval			
Change in Acres %	Corn	Soy	Wheat
Heartland	3.69%	-3.83%	-6.38%
	(1.90% , 12.75%)	(-21.85% , -1.52%)	(-25.13% , 97.45%)
Northern Crescent	2.11%	-3.78%	-1.83%
	(0.412% , 9.285%)	(-18.10% , -1.12%)	(-8.21% , 15.60%)
Northern Great Plains	6.87%	-5.24%	-1.19%
	(2.77% , 22.42%)	(-24.60% , -1.58%)	(-2.51% , 4.24%)
Prairie Gateway	6.60%	-10.28%	-0.31%
	(3.02 , 21.37)	(-57.30 , -2.97)	(-1.671 , 5.044)
Aggregated	4.13%	-4.57%	-1.31%
	(2.37% , 13.70%)	(-24.37% , -2.05%)	(-3.15% , 11.10%)
Price	$5.24	$11.30	$7.09
	($4.10 , $5.39)	($1.97 , $11.92)	($5.09 , $7.33)

Table 3.7: Regional Model Change in Prices and Acres from 20% decrease in Heartland Yields Median & 95% Cred. Interval			
Change in Acres %	Corn	Soy	Wheat
Heartland	-1.35%	0.64%	-1.78%
	(-2.24% , 5.20%)	(-12.98% , 2.17%)	(-15.29% , 82.15%)
Northern Crescent	1.96%	0.76%	-4.36%
	(0.711% , 8.93%)	(-14.46% , 2.79%)	(-14.48% , 13.25%)
Northern Great Plains	3.28%	0.49%	-0.74%
	(0.29% , 18.37%)	(-20.34% , 3.83%)	(-2.47% , 4.90%)
Prairie Gateway	2.66%	0.83%	-1.33%
	(-0.14% , 16.83%)	(-46.18% , 7.74%)	(-2.88% , 4.28%)
Aggregated	0.07%	0.72%	-1.53%
	(-1.02% , 8.21%)	(-16.80% , 2.37%)	(-2.97% , 10.16%)
Price	$5.25	$13.63	$7.23
	($4.15 , $5.35)	($3.63 , $14.65)	($5.45 , $7.43)

APPENDIX

R CODE USED

#Code for national model

#initial setup and loading packages
```
rm(list=ls())
library(MASS)
library(Matrix)
library(matrixcalc)
library(msm)
library(mcmc)
```

#setting time, crops used, iterations, number of chains and loading data found by ERS and NASS
```
t=5
crops=3
iterations=500000
times=4
data=read.csv("US data.csv")
trenddata=read.csv("UScostandyields.csv")
```

#formatting data loading into specific variables of acreage, cost, revenue, price, and total acerage
```
v=array(dim=c(t));
obx=matrix(c(data$Corn.Acres[1:t], data$Soy.Acres[1:t],
data$Wheat.Acres[1:t]),nrow=t,ncol=crops)
v=t(data$Corn.Acres[1:t])+t(data$Soy.Acres[1:t])+t(data$Wheat.Acres[1:t])
cost=matrix(c(data$Corn.Cost.Operations.ERS[1:t], data$Soy.Cost.Operations.ERS[1:t],
data$Wheat.Cost.Operations.ERS[1:t]),nrow=t,ncol=crops)
revenue=matrix(c(data$Corn.Revenue.ERS[1:t], data$Soy.Revenue.ERS[1:t],
data$Wheat.Revenue.ERS[1:t]),nrow=t,ncol=crops)
price=matrix(c(data$Corn.Price.ERS[1:t], data$Soy.Price.ERS[1:t],
data$Wheat.Price.ERS[1:t]),nrow=t,ncol=crops)
#price index used
q=c(232.957, 229.594, 224.939, 218.056, 214.537, 215.303, 207.3, 201.6, 195.3)
expectedprice=matrix(c(data$Corn.Expect.Price[1:t], data$Soy.Expect.Price[1:t],
data$Wheat.Expect.Price[1:t]),nrow=t,ncol=crops)
prevent=matrix(c(data$Corn.Prevented[1:t], data$Soy.Prevented[1:t],
data$Wheat.Prevented[1:t]),nrow=t,ncol=crops)
planted=obx
obx=obx+prevent
```
#linearization and expectations for yield and cost

```
costcorn=trenddata$Cost.Corn[24:34]
costsoy=trenddata$Cost.Soy[24:34]
costwheat=trenddata$Cost.Wheat[24:34]
yieldcorn=trenddata$Corn.Yield[1:34]
yieldsoy=trenddata$Soy.Yield[1:34]
yieldwheat=trenddata$Wheat.Yield[1:34]
trend1=1:11
trend2=1:34
#inputting 2014 acreage, price, and determining total acreage
obx2014=c((90885000+1582408),(84184000+835378),(56822000+1378007))
price2014=c(4.62,11.36,7.02)
v2014=sum(obx2014)
q2014=232.639
#finding the linearization of yield based on time
expected1=lm(yieldcorn~trend2)
expected2=lm(yieldsoy~trend2)
expected3=lm(yieldwheat~trend2)
cost1=lm(costcorn~trend1)
cost2=lm(costsoy~trend1)
cost3=lm(costwheat~trend1)
part1=(12-t):11
part2=(35-t):34
expectedy1=rev(expected1$coef[1]+expected1$coef[2]*part2)
expectedy2=rev(expected2$coef[1]+expected2$coef[2]*part2)
expectedy3=rev(expected3$coef[1]+expected3$coef[2]*part2)
expectedyeild=matrix(c(t(expectedy1),t(expectedy2),t(expectedy3)),nrow=t,ncol=crops)
corny2014=expected1$coef[1]+expected1$coef[2]*35
soyy2014=expected2$coef[1]+expected2$coef[2]*35
wheaty2014=expected3$coef[1]+expected3$coef[2]*35
cornc2014=cost1$coef[1]+cost1$coef[2]*12
soyc2014=cost2$coef[1]+cost2$coef[2]*12
wheatc2014=cost3$coef[1]+cost3$coef[2]*12
expectedyeild2014=c(corny2014,soyy2014,wheaty2014)
cost2014=c(cornc2014,soyc2014,wheatc2014)
revenue=expectedprice*expectedyeild
revenue2014=price2014*expectedyeild2014

#forming diagonal matrices for the elasticity portion of the model
revenuemat=array(dim=c(crops,crops,t))
obxmat=array(dim=c(crops,crops,t))
for(k in 1:t){
        obxmat[,,k]=diag(obx[k,])
        revenuemat[,,k]=diag(revenue[k,])
        v[k]=v[k]+sum(prevent[k,])
}
```

```
revenuemat2014=diag(revenue2014)
totalland=rep(0,t)
for(k in 1:t){
        totalland[k]=sum(v[k])
}
#forming land index
l=rep(0,t)
for(k in 1:t){
        l[k]=(totalland[ceiling(t/2)]/totalland[k])^2
}
l2014=(totalland[ceiling(t/2)]/v2014)^2
qtest=q
R=rep(1,crops)
#finding lambda based off the data
datalambda=rep(0,t)
        for(k in 1:t){
                datalambda[k]=(revenue[k,]-cost[k,])%*%(obx[k,]/v[k])
        }

#finding the point to set U to in the model
a=.25*mean(obx[,1]/(expectedyeild[,1]*expectedprice[,1]))
b=.25*mean(obx[,2]/(expectedyeild[,2]*expectedprice[,2]))
c=.15*mean(obx[,3]/(expectedyeild[,3]*expectedprice[,3]))

                u1=sqrt(-(1/(-2*a+2*b+2*c))*(-2*a*b-2*a*c-2*b*c+a^2+b^2+c^2)*mean(q))
                u2=sqrt(-(1/(2*a-2*b+2*c))*(-2*a*b-2*a*c-2*b*c+a^2+b^2+c^2)*mean(q))
                u3=sqrt(-(1/(2*a+2*b-2*c))*(-2*a*b-2*a*c-2*b*c+a^2+b^2+c^2)*mean(q))


#function to find x(theta) and lambda(theta)
modelfunction <- function(param){
        lambda=array(dim=c(t))
        x=array(dim=c(crops,t))

        xtotal=array(dim=c(crops*t))
        elasticitly=array(dim=c(crops*crops,t))
        E=matrix(0,nrow=crops,ncol=crops)
        count=1
        for(k in 1:crops){
                E[k,k]=(param[count])*(param[count])
                count=count+1
                }
        c=param[count:(count+crops-1)];
```

```
                for(k in 1:t){
                        inverseE=1/l[k]*1/q[k]*E;
                        lambda[k]=(1/(t(R)%*%inverseE%*%(R)))
                        *(t(R)%*%inverseE%*%(revenue[k,]-cost[k,]-qtest[k]*c)-v[k]);
                        x[,k]=inverseE%*%(revenue[k,]-cost[k,]-qtest[k]*c-R*lambda[k]);
                }
        return(c(lambda,t(vec(x))))
}




# likelihood function based on theta (param)
like <- function(param){
        dlambda=rep(0,t);
        dx=array(dim=c(crops,t))

        for (k in 1:t){
                        lambdamean=param[k]
                        xmean=(param[(t+1+crops*(k-1)):(t+crops+crops*(k-1))])
                        dlambda[k]=dtnorm(datalambda[k],lambdamean,(.25
                        *abs(datalambda[k])),lower=0, log=T);
                        if(lambdamean<=0)dlambda[k]=-Inf
                        for(j in 1:crops){
                        dx[j,k]=dtnorm((obx[k,j]),xmean[j],(.2/3*abs(obx[k,j])),lower=0, log=T)
                        if(xmean[j]<=0)dx[j,k]=-Inf
                        }
        }
    li=sum(dlambda)+sum(dx);
    return(li)
}




#function to find prior density based on theta (param)
prior <- function(param){
        dparam=rep(0,crops*2)
        dparam[1]=dnorm(param[1],(u1),700, log=T)
        dparam[2]=dnorm(param[2],(u2),700, log=T)
        dparam[3]=dnorm(param[3],(u3),700, log=T)
        mean1=-1*(mean(obx[,1])/(param[1]^2/mean(q))-(mean(revenue[,1])-mean(cost[,1])-
        mean(datalambda)))/mean(q)
```

```
        mean2=-1*(mean(obx[,2])/(param[2]^2/mean(q))-(mean(revenue[,2])-mean(cost[,2])-
        mean(datalambda)))/mean(q)
        mean3=-1*(mean(obx[,3])/(param[3]^2/mean(q))-(mean(revenue[,3])-mean(cost[,3])-
        mean(datalambda)))/mean(q)

        dparam[4]=dnorm(param[4],mean1,5, log=T)
        dparam[5]=dnorm(param[5],mean2,5, log=T)
        dparam[6]=dnorm(param[6],mean3,5, log=T)

    pri=sum(dparam)
    return(pri)
}

#function to define the posterior density for a given theta
posterior <- function(param){
        value=modelfunction(param)
    return (like(value)+prior(param))
}

#proposal function to draw a new theta
proposalfunction <- function(param){
                modelu=rtnorm(3,mean=c(param[1],param[2],param[3]),sd=rep(1000,crops))
                meanc1=(mean(obx[,1])/(modelu[1]^2/mean(q))-(mean(revenue[,1])-
                mean(cost[,1])-mean(datalambda)))/mean(q)
                meanc2=(mean(obx[,2])/(modelu[2]^2/mean(q))-(mean(revenue[,2])-
                mean(cost[,2])-mean(datalambda)))/mean(q)
                meanc3=(mean(obx[,3])/(modelu[3]^2/mean(q))-(mean(revenue[,3])-
                mean(cost[,3])-mean(datalambda)))/mean(q)
                modelc=rnorm(crops,mean = c((-1*meanc1),(-1*meanc2),(-1*meanc3)),
                sd=rep(1,crops))
                model=c(modelu,modelc);
                return(model);
}


#body of the  MCMC algorithm
run _MCMC <- function(startvalue, iterations){
    chain = array(dim = c(iterations+1,length(startvalue)))
    chain[1,] = startvalue
    for (i in 1:iterations){
        proposal = proposalfunction(chain[i,])
        probab = (posterior(proposal) - posterior(chain[i,]))
            u=log(runif(1))
        if (u < probab){
```

```
        chain[i+1,] = proposal
      }else{
        chain[i+1,] = chain[i,]
      }
   }
   return(chain)
}

#function to create a random starting point based on alternative elasticity
testC=array(dim=c(crops,t))
testxd=array(dim=c(t,crops))
startmodel=array(dim=c(times,(crops+crops)))
Q=diag(crops)*1
for(i in 1:times){
        #test to determining if starting value is valid
        test=-Inf
        while(test==-Inf){
                startu=rtnorm(3,3000,sd=rep(10000,3))
                mean1=-1*(mean(obx[,1])/(startu[1]^2/mean(q))-(mean(revenue[,1])-
                mean(cost[,1])-mean(datalambda)))/mean(q)
                mean2=-1*(mean(obx[,2])/(startu[2]^2/mean(q))-(mean(revenue[,2])-
                mean(cost[,2])-mean(datalambda)))/mean(q)
                mean3=-1*(mean(obx[,3])/(startu[3]^2/mean(q))-(mean(revenue[,3])-
                mean(cost[,3])-mean(datalambda)))/mean(q)
                startc=rnorm(3,mean=c(mean1,mean2,mean3),sd=rep(5,3))
                startmodel[i,]=c(startu,startc)
                test=posterior(startmodel[i,])
        }
}

acceptencerate=array(dim=c(times));

#function to look several chains
        for(m in 1:times){
        chain = run_MCMC(startmodel[m,], iterations)
        #test of the acceptance rate
        acceptencerate[m]=(1-mean(duplicated(chain)))*100
        name1=paste("National model+",m,sep="_")
        save(chain, file=name1)
        startmodel[m,]=chain[iterations+1,]
        rm(chain)
        }
```

```
#regional model clearing data and initalizing
rm(list=ls())
library(MASS)
library(Matrix)
library(matrixcalc)
library(msm)
library(mcmc)
library(DirichletReg)


#setting time, crops used, iterations, number of chains and loading data
t=5
crops=3
r=4
iterations=500000
times=6
load("regionalacres.RData")
obx=acres[(2:5),,(1:3),1]+acres[(2:5),,(1:3),2]
revenue=array(dim=c(r,t,crops))
expectedyeild=array(dim=c(r,t,crops))
cost=array(dim=c(r,t,crops))
trend=(c(1:(34)))
cornyield=read.csv("cornyeilds.csv")
soyyield=read.csv("soyyeilds.csv")
wheatyield=read.csv("wheatyeilds.csv")
corndata=matrix(c(cornyield$Iowa, cornyield$Michagan, cornyield$North.Dakota,
cornyield$Kansas),nrow=34,ncol=4)
soydata=matrix(c(soyyield$Iowa, soyyield$Michagan, soyyield$North.Dakota,
soyyield$Kansas),nrow=34,ncol=4)
wheatdata=matrix(c(wheatyield$Iowa,wheatyield$Michagan, wheatyield$North.Dakota,
wheatyield$Kansas),nrow=34,ncol=4)
start=9-t+1

#loading in the regional acreage yield, price, expected price and yeild
for(i in 1:r){
if(i==1)data=read.csv("HeartLand.csv")
if(i==2)data=read.csv("Northwest Cresent.csv")
if(i==3)data=read.csv("Northern Great Plains.csv")
if(i==4)data=read.csv("Pairie Gateway.csv")

corncost=data$Corn.Cost[start:9]
cornexpectprice=data$Corn.Expected.Price[start:9]
added=data$Corn.Add.Value[start:9]
```

```
yieldlin1=lm(corndata[,i]~trend)
expectedyeild[i,,1]=(yieldlin1$coef[1]+yieldlin1$coef[2]*trend[(35-t):34])
revenue[i,,1]=(yieldlin1$coef[1]+yieldlin1$coef[2]*trend[(35-t):34])*cornexpectprice+added
cost[i,,1]=corncost

soycost=data$Soy.Cost[start:9]
soyexpectprice=data$Soy.Expected.Price[start:9]
yieldlin2=lm(soydata[,i]~trend)
expectedyeild[i,,2]=(yieldlin2$coef[1]+yieldlin2$coef[2]*trend[(35-t):34])
revenue[i,,2]=(yieldlin2$coef[1]+yieldlin2$coef[2]*trend[(35-t):34])*soyexpectprice
cost[i,,2]=soycost




wheatcost=data$Wheat.Cost[start:9]
wheatexpectprice=data$Wheat.Expected.Price[start:9]
added=data$Wheat.Added.Value[start:9]
yieldlin3=lm(wheatdata[,i]~trend)
expectedyeild[i,,3]=(yieldlin3$coef[1]+yieldlin3$coef[2]*trend[(35-t):34])
revenue[i,,3]=(yieldlin3$coef[1]+yieldlin3$coef[2]*trend[(35-t):34])*wheatexpectprice+added
cost[i,,3]=wheatcost
}

v=array(dim=c(r,t))
revenuemat=array(dim=c(crops,crops,r,t))
obxmat=array(dim=c(crops,crops,r,t))
yieldmat=array(dim=c(crops,crops,r,t))
for(i in 1:r){
        for(j in 1:t){
                v[i,j]=sum(obx[i,j,])
                obxmat[,,i,j]=diag(obx[i,j,])
                revenuemat[,,i,j]=diag(revenue[i,j,])
                yieldmat[,,i,j]=diag(expectedyeild[i,j,])
        }
}


#calculating l index
l=array(dim=c(r,t))
for(i in 1:r){
        for(j in 1:t){
                l[i,j]=(v[i,ceiling(t/2)]/v[i,j])^2
        }
}
```

```
#loading in price index
q=rev(c(232.957, 229.594, 224.939, 218.056, 214.537))

R=rep(1,crops)
datalambda=array(dim=c(r,t))
for(i in 1:r){
        for(j in 1:t){
                datalambda[i,j]=(revenue[i,j,]-cost[i,j,])%*%(obx[i,j,]/v[i,j])
        }
}
L
```

```
#mean of the various slopes based on the posterior density of the national
nationalcornslope=4630505
nationalsoyslope=1676530
nationalwheatslope=1122711
data=read.csv("US data.csv")
obn=matrix(c(rev(data$Corn.Acres[1:t]), rev(data$Soy.Acres[1:t]),
        rev(data$Wheat.Acres[1:t])),nrow=t,ncol=crops)
prevent=matrix(c(rev(data$Corn.Prevented[1:t]), rev(data$Soy.Prevented[1:t]),
        rev(data$Wheat.Prevented[1:t])),nrow=t,ncol=crops)
national=obn+prevent
Hearttotal=c(sum(obx[1,1,]),sum(obx[1,2,]),sum(obx[1,3,]),sum(obx[1,4,]),sum(obx[1,5,]))
NCtotal=c(sum(obx[2,1,]),sum(obx[2,2,]),sum(obx[2,3,]),sum(obx[2,4,]),sum(obx[2,5,]))
NGPtotal=c(sum(obx[3,1,]),sum(obx[3,2,]),sum(obx[3,3,]),sum(obx[3,4,]),sum(obx[3,5,]))
Pairietotal=c(sum(obx[4,1,]),sum(obx[4,2,]),sum(obx[4,3,]),sum(obx[4,4,]),sum(obx[4,5,]))
Nationaltotal=c(sum(national[1,]),sum(national[2,]),sum(national[3,]),sum(national[4,]),sum(nati
        onal[5,]))
#determine the land percentage
percent1=mean(Hearttotal/Nationaltotal)
percent2=mean(NCtotal/Nationaltotal)
percent3=mean(NGPtotal/Nationaltotal)
percent4=mean(Pairietotal/Nationaltotal)
percent5=1-percent1-percent2-percent3-percent4
landpercent=c(percent1,percent2,percent3,percent4,percent5)
regionslicecorn=c(0,0,0,0)
regionslicesoy=c(0,0,0,0)
regionslicewheat=c(0,0,0,0)
for( spot in 1:4){
regionslicecorn[spot]=nationalcornslope/mean(expectedyeild[spot,,1])
regionslicesoy[spot]=nationalsoyslope/mean(expectedyeild[spot,,2])
regionslicewheat[spot]=nationalwheatslope/mean(expectedyeild[spot,,3])
```

```
}


#load the slope found by the posterior density
load("National slope")
slopetest=slopetotal
rm(slopetotal)
slopetotal=array(dim=c(length(slopetest[,1])/100,3))
for(i in 1:(length(slopetest[,1])/100)){
        slopetotal[i,1]=mean(slopetest[i*100,1],slopetest[i*100,10],slopetest[i*100,19],slopetest[
        i*100,28],slopetest[i*100,37])
        slopetotal[i,2]=mean(slopetest[i*100,5],slopetest[i*100,14],slopetest[i*100,23],slopetest[
        i*100,32],slopetest[i*100,41])
        slopetotal[i,3]=mean(slopetest[i*100,9],slopetest[i*100,18],slopetest[i*100,27],slopetest[
        i*100,36],slopetest[i*100,45])
}


#function to find x(r,theta) and lamda(r,theta)
modelfunction <- function(m,param){
        lambda=array(dim=c(t))
        x=array(dim=c(crops,t))

        xtotal=array(dim=c(crops*t))
        elasticitly=array(dim=c(crops*crops,t))
        E=matrix(0,nrow=crops,ncol=crops)
        count=1
        for(k in 1:crops){
                E[k,k]=(param[count])*(param[count])
                count=count+1
                }
        c=param[count:(count+crops-1)];
                for(k in 1:t){
                inverseE=1/l[m,k]*1/qtest[k]*E;
                lambda[k]=(1/(t(R)%*%inverseE%*%(R)))*(t(R)%*%inverseE%*%(revenue[m,
                k,]-cost[m,k,]-qtest[k]*c)-v[m,k]);
                x[,k]=inverseE%*%(revenue[m,k,]-cost[m,k,]-qtest[k]*c-R*lambda[k]);
                }
        return(c(lambda,t(vec(x))))
}
```

```
#function to find likelihood denisty based on theta (param) and region r
like <- function(m,param){
        dlambda=rep(0,t);
        dx=array(dim=c(crops,t))


        for (k in 1:t){

                        lambdamean=param[k]
                        xmean=(param[(t+1+crops*(k-1)):(t+crops+crops*(k-1))])

                dlambda[k]=dtnorm(datalambda[m,k],lambdamean,(abs(datalambda[m,k])),lower
        =0, log=T);
                        if(lambdamean<=0)dlambda[k]=-Inf
                        for(j in 1:crops){
                        dx[j,k]=dtnorm((obx[m,k,j]),xmean[j],(abs(obx[m,k,j])),lower=0, log=T)
                        if(xmean[j]<=0)dx[j,k]=-Inf
                        }

        }

   li=sum(dlambda)+sum(dx);
   return(li)
}

#function to find prior density for U and l based on theta (param) and region r
prior <- function(m,param){
        dparam=rep(0,crops*3)
        pull=param[7]
        nationalcornslope=slopetotal[pull,1]
        nationalsoyslope=slopetotal[pull,2]
        nationalwheatslope=slopetotal[pull,3]
        a=nationalcornslope/mean(expectedyeild[m,,1])*(param[8])
        b=nationalsoyslope/mean(expectedyeild[m,,2])*(param[9])
        c=nationalwheatslope/mean(expectedyeild[m,,3])*(param[10])
        u1=-(1/(-2*a+2*b+2*c))*(-2*a*b-2*a*c-2*b*c+a^2+b^2+c^2)*mean(q)
        u2=-(1/(2*a-2*b+2*c))*(-2*a*b-2*a*c-2*b*c+a^2+b^2+c^2)*mean(q)
        u3=-(1/(2*a+2*b-2*c))*(-2*a*b-2*a*c-2*b*c+a^2+b^2+c^2)*mean(q)
        meanc1=-(mean(obx[m,,1])/(param[1]^2/mean(q))-(mean(revenue[m,,1]-
        mean(cost[m,,1])-mean(datalambda[m,])))/mean(q)
        meanc2=-(mean(obx[m,,2])/(param[2]^2/mean(q))-(mean(revenue[m,,2])-
        mean(cost[m,,2])-mean(datalambda[m,])))/mean(q)
```

```
        meanc3=-(mean(obx[m,,3])/(param[3]^2/mean(q))-(mean(revenue[m,,3])-
        mean(cost[m,,3])-mean(datalambda[m,])))/mean(q)

        dparam[1]=dnorm(param[1],sqrt(u1),50, log=T)
        dparam[2]=dnorm(param[2],sqrt(u2),50, log=T)
        dparam[3]=dnorm(param[3],sqrt(u3),50, log=T)
        dparam[4]=dnorm(param[4],meanc1,.5, log=T)
        dparam[5]=dnorm(param[5],meanc2,.5, log=T)
        dparam[6]=dnorm(param[6],meanc3,.5, log=T)

        pri=sum(dparam)
        return(pri)
}



#function to find prior density alpha based on theta (param) and region r
prior2 <- function(param){
        dparam=rep(0,3)
        cornextra=1-sum(param[1,])
        soyextra=1-sum(param[2,])
        wheatextra=1-sum(param[3,])
        alpha1=matrix(c(param[1,],cornextra),nrow=1,ncol=5)
        alpha2=matrix(c(param[2,],soyextra),nrow=1,ncol=5)
        alpha3=matrix(c(param[3,],wheatextra),nrow=1,ncol=5)

        dparam[1]=ddirichlet(alpha1,landpercent*100,log=T)
        dparam[2]=ddirichlet(alpha2,landpercent*100,log=T)
        dparam[3]=ddirichlet(alpha3,landpercent*100,log=T)

         pri=sum(dparam)
        return(pri)
}

#posterior density combing the prior and likelihood
posterior <- function(param){
        sumofprob=c(0,0,0,0)
        percentvect=array(dim=c(3,r))
        for(m in 1:r){
                value=modelfunction(m,param[m,])
                percentvect[1,m]=param[m,8]
                percentvect[2,m]=param[m,9]
                percentvect[3,m]=param[m,10]
                sumofprob[m]=like(m,value)+prior(m,param[m,])
        }
```

```
        portprior=prior2(percentvect)
        return (sum(sumofprob)+portprior)
}


#function to determine the probability from going from param1 to param2
proposaljump <- function(param1,param2){
        dparam1=rep(0,3)
        cornextra=1-sum(param1[,1])
        soyextra=1-sum(param1[,2])
        wheatextra=1-sum(param1[,3])
        alpha11=matrix(c(param1[,1],cornextra),nrow=1,ncol=5)
        alpha21=matrix(c(param1[,2],soyextra),nrow=1,ncol=5)
        alpha31=matrix(c(param1[,3],wheatextra),nrow=1,ncol=5)

        dparam2=rep(0,3)
        cornextra=1-sum(param2[,1])
        soyextra=1-sum(param2[,2])
        wheatextra=1-sum(param2[,3])
        alpha12=matrix(c(param2[,1],cornextra),nrow=1,ncol=5)
        alpha22=matrix(c(param2[,2],soyextra),nrow=1,ncol=5)
        alpha32=matrix(c(param2[,3],wheatextra),nrow=1,ncol=5)

        dparam1[1]=ddirichlet(alpha11,alpha12*200,log=T)
        dparam1[2]=ddirichlet(alpha21,alpha22*200,log=T)
        dparam1[3]=ddirichlet(alpha31,alpha32*200,log=T)

        dparam2[1]=ddirichlet(alpha12,alpha11*200,log=T)
        dparam2[2]=ddirichlet(alpha22,alpha21*200,log=T)
        dparam2[3]=ddirichlet(alpha32,alpha31*200,log=T)

         pri=sum(dparam1)-sum(dparam2)
         return(pri)
}

#function to draw a new alpha, U, and l based off of current theta and the current theta
#and posterior density of national slope
proposalfunction <- function(param){
             check=0
             landratio=array(dim=c(crops,r+1))
             cornextra=max(1-sum(param[,8]),0.001)
             soyextra=max(1-sum(param[,9]),0.001)
             wheatextra=max(1-sum(param[,10]),0.001)
             alpha1=matrix(c(param[,8],cornextra),nrow=1,ncol=5)
             alpha2=matrix(c(param[,9],soyextra),nrow=1,ncol=5)
             alpha3=matrix(c(param[,10],wheatextra),nrow=1,ncol=5)
```

```
                while(check==0){
                pull=ceiling(18000*runif(1))
                nationalcornslope=slopetotal[pull,1]
                nationalsoyslope=slopetotal[pull,2]
                nationalwheatslope=slopetotal[pull,3]

                check=1
                landratio[1,]=rdirichlet(1,alpha1*200)
                landratio[2,]=rdirichlet(1,alpha2*200)
                landratio[3,]=rdirichlet(1,alpha3*200)
                u1=c(0,0,0,0)
                u2=c(0,0,0,0)
                u3=c(0,0,0,0)
                for(m in 1:r){
                a=nationalcornslope/mean(expectedyeild[m,,1])*(landratio[1,m])
                b=nationalsoyslope/mean(expectedyeild[m,,2])*(landratio[2,m])
                c=nationalwheatslope/mean(expectedyeild[m,,3])*(landratio[3,m])
                u1[m]=-(1/(-2*a+2*b+2*c))*(-2*a*b-2*a*c-2*b*c+a^2+b^2+c^2)*mean(q)
                u2[m]=-(1/(2*a-2*b+2*c))*(-2*a*b-2*a*c-2*b*c+a^2+b^2+c^2)*mean(q)
                u3[m]=-(1/(2*a+2*b-2*c))*(-2*a*b-2*a*c-2*b*c+a^2+b^2+c^2)*mean(q)
                }
                if(any(u1<0))check=0
                if(any(u2<0))check=0
                if(any(u3<0))check=0
                }
                model=array(dim=c(crops*3+1,r))
                for(m in 1:r){
                        modelu=rnorm(crops,mean = c(sqrt(u1[m]),sqrt(u2[m]),sqrt(u3[m])),
                        sd=rep(50,crops))
                        meanc1=-(mean(obx[m,,1])/(modelu[1]^2/mean(q))-
                        (mean(revenue[m,,1])-mean(cost[m,,1])-mean(datalambda[m,])))/mean(q)
                        meanc2=-(mean(obx[m,,2])/(modelu[2]^2/mean(q))-
                        (mean(revenue[m,,2])-mean(cost[m,,2])-mean(datalambda[m,])))/mean(q)
                        meanc3=-(mean(obx[m,,3])/(modelu[3]^2/mean(q))-
                        (mean(revenue[m,,3])-mean(cost[m,,3])-mean(datalambda[m,])))/mean(q)
                        modelc=rnorm(crops,mean = c(meanc1,meanc2,meanc3),
                        sd=rep(.5,crops))
                        model[,m]=c(modelu,modelc,pull,landratio[1,m],landratio[2,m],
                        landratio[3,m])
                }
        return(t(model));
}

#function to find a starting point for theta
startmodel= array(dim = c(times,r,crops*3+1))
```

```
startratio=array(dim=c(3,r))
for(m in 1:times){
        scheck=0
        check2=0
                while(check2==0){
                        check2=1
                        landpercent1=rdirichlet(1,landpercent*10)
                        landpercent2=rdirichlet(1,landpercent*10)
                        landpercent3=rdirichlet(1,landpercent*10)

                        spull=ceiling(18000*runif(1))
                        snationalcornslope=slopetotal[spull,1]
                        snationalsoyslope=slopetotal[spull,2]
                        snationalwheatslope=slopetotal[spull,3]
                        su1=c(0,0,0,0)
                        su2=c(0,0,0,0)
                        su3=c(0,0,0,0)
                        for(j in 1:r){
                                sa=snationalcornslope/mean(expectedyeild[j,,1])*(landpercent1[j])
                                sb=snationalsoyslope/mean(expectedyeild[j,,2])*(landpercent2[j])
                                sc=snationalwheatslope/mean(expectedyeild[j,,3])
                                *(landpercent3[j])
                                su1[j]=-(1/(-2*sa+2*sb+2*sc))*(-2*sa*sb-2*sa*sc-
                                2*sb*sc+sa^2+sb^2+sc^2)*mean(q)
                                su2[j]=-(1/(2*sa-2*sb+2*sc))*(-2*sa*sb-2*sa*sc-
                                2*sb*sc+sa^2+sb^2+sc^2)*mean(q)
                                su3[j]=-(1/(2*sa+2*sb-2*sc))*(-2*sa*sb-2*sa*sc-
                                2*sb*sc+sa^2+sb^2+sc^2)*mean(q)
                        }
                        if(any(su1<0))check2=0
                        if(any(su2<0))check2=0
                        if(any(su3<0))check2=0
                }
        while(scheck==0){
                scheck=1
                        start1=array(dim=c(crops*3+1,r))

                for(i in 1:r){
                        start1[,i]=c(0,0,0,0,0,0,0,landpercent1[i],landpercent2[i],landpercent3[i])
                }
                test=proposalfunction(t(start1))
                if(posterior(test)==-Inf)scheck=0
        }
        startmodel[m,,]=test
}
```

```
run_MCMC <- function(startvalue, iterations){
    chain = array(dim = c(iterations+1,r,crops*3+1))
    chain[1,,] = startvalue
    for (i in 1:iterations){
        proposal = proposalfunction(chain[i,,])
            jumpratio=proposaljump(proposal[,8:10],chain[i,,8:10])
        probab = (posterior(proposal) - posterior(chain[i,,]))+jumpratio

            u=log(runif(1))
        if (u < probab){
            chain[i+1,,] = proposal
        }else{
            chain[i+1,,] = chain[i,,]
        }
    }
    return(chain)
}

#function to look several chains
acceptencerate=array(dim=c(times));
for(m in 1:times){
        chain = run_metropolis_MCMC(startmodel[m,,], iterations)
        acceptencerate[m]=(1-mean(duplicated(chain)))*100
        name1=paste("Regional",m,sep="_")
        save(chain, file=name1)
        startmodel[m,,]=chain[iterations+1,,]
        rm(chain)

}
```

```
#code to determine regional shift
#clearing data and loading library
rm(list=ls())
library(MASS)
library(Matrix)
library(matrixcalc)
library(msm)
library(mcmc)
library(DirichletReg)

#setting time, crops used, iterations, number of chains and loading data
t=5
crops=3
r=4

#loading data for acerage cost and yields includeing 2014 information to test against
load("regionalacresplus2014.RData")
obx=acres[(2:5),,(1:3),1]+acres[(2:5),,(1:3),2]
revenue=array(dim=c(r,t,crops))
expectedyeild=array(dim=c(r,t,crops))
cost=array(dim=c(r,t,crops))
cost2014=array(dim=c(r,crops))
revenue2014=array(dim=c(r,crops))
expectedyeild2014=array(dim=c(r,crops))
expectedprice2014=c(4.62,11.36,7.02)
trend=(c(1:(34)))
cornyield=read.csv("cornyeilds.csv")
soyyield=read.csv("soyyeilds.csv")
wheatyield=read.csv("wheatyeilds.csv")
corndata=matrix(c(cornyield$Iowa, cornyield$Michagan, cornyield$North.Dakota,
cornyield$Kansas),nrow=34,ncol=4)
soydata=matrix(c(soyyield$Iowa, soyyield$Michagan, soyyield$North.Dakota,
soyyield$Kansas),nrow=34,ncol=4)
wheatdata=matrix(c(wheatyield$Iowa,wheatyield$Michagan, wheatyield$North.Dakota,
wheatyield$Kansas),nrow=34,ncol=4)
start=9-t+1
trend2=c(1:t)

#loading data into matrices to use for predictions and shifts
for(i in 1:r){
if(i==1)data=read.csv("HeartLand.csv")
if(i==2)data=read.csv("Northwest Cresent.csv")
if(i==3)data=read.csv("Northern Great Plains.csv")
if(i==4)data=read.csv("Pairie Gateway.csv")
```

```
corncost=data$Corn.Cost[start:9]
cornexpectprice=data$Corn.Expected.Price[start:9]
added=data$Corn.Add.Value[start:9]
yieldlin1=lm(corndata[,i]~trend)
costlin1=lm(data$Corn.Cost[start:9]~trend2)
expectedyeild[i,,1]=(yieldlin1$coef[1]+yieldlin1$coef[2]*trend[(35-t):34])
revenue[i,,1]=(yieldlin1$coef[1]+yieldlin1$coef[2]*trend[(35-t):34])*cornexpectprice+added
cost[i,,1]=corncost
cost2014[i,1]=(costlin1$coef[1]+costlin1$coef[2]*6)
expectedyeild2014[i,1]=(yieldlin1$coef[1]+yieldlin1$coef[2]*35)
revenue2014[i,1]=(yieldlin1$coef[1]+yieldlin1$coef[2]*35)*expectedprice2014[1]

soycost=data$Soy.Cost[start:9]
soyexpectprice=data$Soy.Expected.Price[start:9]
yieldlin2=lm(soydata[,i]~trend)
costlin2=lm(data$Soy.Cost[start:9]~trend2)
expectedyeild[i,,2]=(yieldlin2$coef[1]+yieldlin2$coef[2]*trend[(35-t):34])
revenue[i,,2]=(yieldlin2$coef[1]+yieldlin2$coef[2]*trend[(35-t):34])*soyexpectprice
cost[i,,2]=soycost
cost2014[i,2]=(costlin2$coef[1]+costlin2$coef[2]*6)
expectedyeild2014[i,2]=(yieldlin2$coef[1]+yieldlin2$coef[2]*35)
revenue2014[i,2]=(yieldlin2$coef[1]+yieldlin2$coef[2]*35)*expectedprice2014[2]

wheatcost=data$Wheat.Cost[start:9]
wheatexpectprice=data$Wheat.Expected.Price[start:9]
added=data$Wheat.Added.Value[start:9]
yieldlin3=lm(wheatdata[,i]~trend)
costlin1=lm(data$Wheat.Cost[start:9]~trend2)
expectedyeild[i,,3]=(yieldlin3$coef[1]+yieldlin3$coef[2]*trend[(35-t):34])
revenue[i,,3]=(yieldlin3$coef[1]+yieldlin3$coef[2]*trend[(35-t):34])*wheatexpectprice+added
cost[i,,3]=wheatcost
cost2014[i,3]=(costlin2$coef[1]+costlin2$coef[2]*6)
expectedyeild2014[i,3]=(yieldlin3$coef[1]+yieldlin3$coef[2]*35)
revenue2014[i,3]=(yieldlin3$coef[1]+yieldlin3$coef[2]*35)*expectedprice2014[3]
}
v=array(dim=c(r,t))
v2014=array(dim=c(r))
revenuemat=array(dim=c(crops,crops,r,t))
revenuemat2014=array(dim=c(crops,crops,r))
obxmat=array(dim=c(crops,crops,r,t))
yieldmat=array(dim=c(crops,crops,r,t))
for(i in 1:r){
        v2014[i]=sum(obx[i,6,])
        revenuemat2014[,,i]=diag(revenue2014[i,])
        for(j in 1:t){
```

```
                v[i,j]=sum(obx[i,j,])
                obxmat[,,i,j]=diag(obx[i,j,])
                revenuemat[,,i,j]=diag(revenue[i,j,])
                yieldmat[,,i,j]=diag(expectedyeild[i,j,])
        }
}

l=array(dim=c(r,t))
l2014=array(dim=c(r))
for(i in 1:r){
        l2014[i]=(v[i,ceiling(t/2)]/v2014[i])^2
        for(j in 1:t){
                l[i,j]=(v[i,ceiling(t/2)]/v[i,j])^2
        }
}

q=rev(c(232.957, 229.594, 224.939, 218.056, 214.537))
q2014=232.639
R=rep(1,crops)
datalambda=array(dim=c(r,t))
datalambda2014=array(dim=c(r))
for(i in 1:r){
        datalambda2014[i]=(revenue2014[i,]-cost2014[i,])%*%(obx[i,6,]/v2014[i])
        for(j in 1:t){
                datalambda[i,j]=(revenue[i,j,]-cost[i,j,])%*%(obx[i,j,]/v[i,j])
        }
}

#determine the linearized function for how lambda and total land relates.
lambda=0
region1forcast=lm(c(v[1,],v2014[1])~c(datalambda[1,],datalambda2014[1]))
region2forcast=lm(c(v[2,],v2014[2])~c(datalambda[2,],datalambda2014[2]))
region3forcast=lm(c(v[3,],v2014[3])~c(datalambda[3,],datalambda2014[3]))
region4forcast=lm(c(v[4,],v2014[4])~c(datalambda[4,],datalambda2014[4]))
region1function=93489888+5448*lambda
region2function=17353569+4467*lambda
region3function=30480680+16254*lambda
region4function=44273522-14926*lambda
total=185597659+11243*lambda

#projecting 2014 data
datalambda2014test=array(dim=c(r))
for(m in 1:r){
```

```
datalambda2014test[m]=(revenue2014[m,]-cost2014[m,])%*%(obx[m,5,]/v[m,5])
if(m==1)slope=region1forcast$coef[2]
if(m==2)slope=region2forcast$coef[2]
if(m==3)slope=region3forcast$coef[2]
if(m==4)slope=0
v2014[m]=v[m,5]+slope*(datalambda2014test[m]-datalambda[m,5])
}
l2014=array(dim=c(r))
for(i in 1:r){
        l2014[i]=(v[i,ceiling(t/2)]/v2014[i])^2
}

#function to find elasticity(theata,r), x(theta,r), and lambda(theta,r)
modelfunction <- function(m,param){
        lambda=array(dim=c(t))
        x=array(dim=c(crops,t))

        xtotal=array(dim=c(crops*t))
        elasticitly=array(dim=c(crops*crops,t))
        E=matrix(0,nrow=crops,ncol=crops)
        count=1
        for(k in 1:crops){
                E[k,k]=(param[count])*(param[count])
                count=count+1
                }
        c=param[count:(count+crops-1)];
        xmat=array(dim=c(crops,crops,t))

                for(k in 1:t){
                        inverseE=1/l[m,k]*1/qtest[k]*E;
                        lambda[k]=(1/(t(R)%*%inverseE%*%(R)))*
                        (t(R)%*%inverseE%*%(revenue[m,k,]-cost[m,k,]-qtest[k]*c)-v[m,k]);
                        x[,k]=inverseE%*%(revenue[m,k,]-cost[m,k,]-qtest[k]*c-R*lambda[k]);
                xmat[,,k]=diag(x[,k]);
                middel=ginv(xmat[,,k])%*%(inverse-inverseE%*%(R)%
                *%ginv(t(R)%*%inverseE%*%(R))%*%t(R)%*%inverseE)%
                *%revenuemat[,,m,k];
                elasticitly[,k]=t(vec(middel))
                }


        return(c(lambda,t(vec(x)),t(vec(elasticitly))))
}

#function to determing the future value of x and theta for 2014
```

```
modelfunctionfuture <- function(m,param){

        x=array(dim=c(crops,t))
        xtotal=array(dim=c(crops*t))
        elasticitly=array(dim=c(crops*crops,t))
        E=matrix(0,nrow=crops,ncol=crops)
        count=1
        for(k in 1:crops){
                E[k,k]=(param[count])*(param[count])
                count=count+1
                }
        c=param[count:(count+crops-1)];
        xmat=array(dim=c(crops,crops,t));
        inverseE=1/l2014[m]*1/q2014*E;
        lambda=(1/(t(R)%*%inverseE%*%(R)))*(t(R)%*%inverseE%*%(revenue2014[m,]-
        cost2014[m,]-q2014*c)-v2014[m]);
        x[,1]=inverseE%*%(revenue2014[m,]-cost2014[m,]-q2014*c-R*lambda);
        return(c(lambda,t(x[,1])))
}

#function used to determine the acreage allocated given a change in in expected profits
modelfunctionfuturechanged <- function(m,i,param){
        x=array(dim=c(crops,t))
        xtotal=array(dim=c(crops*t))
        elasticitly=array(dim=c(crops*crops,t))
        E=matrix(0,nrow=crops,ncol=crops)
        count=1
        for(k in 1:crops){
                E[k,k]=(param[count])*(param[count])
                count=count+1
                }




        c=param[count:(count+crops-1)];
        xmat=array(dim=c(crops,crops,t));
        inverseE=1/l2014changed[m,i]*1/q2014*E;
        lambda=(1/(t(R)%*%inverseE%*%(R)))*(t(R)%*%inverseE%*%(predictedrevenue[m,,i
        ]-cost2014[m,]-q2014*c)-totalpredictedland2014[m,i]);
        x[,1]=inverseE%*%(predictedrevenue[m,,i]-cost2014[m,]-q2014*c-R*lambda);

        return(c(lambda,t(x[,1])))
}
```

```
size=(500000-100000)/10*6

load("2014 regional predictions")
load("Regional Posterior density")
name1="Regional result"


#setup for change, setting the number of iterations it will cycle through supply and demand
duration=25
prediction=predicttotal
predict=modelfunctionfuture(1,chaintotal[1,1,])
#seting up the intial price change
predictionchanged=array(dim=c(size,r,length(predict),duration))
pricechanged=array(dim=c(3,duration+1,size))
for(i in 1:size){
pricechanged[,1,i]=expectedprice2014*c(.8,1,1)
}
#finding the acerage allocated from the prediction
cornland=prediction[,1,2]*expectedyeild2014[1,1]+prediction[,2,2]*expectedyeild2014[2,1]+pre
        diction[,3,2]*expectedyeild2014[3,1]+prediction[,4,2]*expectedyeild2014[4,1]
soyland=prediction[,1,3]*expectedyeild2014[1,2]+prediction[,2,3]*expectedyeild2014[2,2]+pred
        iction[,3,3]*expectedyeild2014[3,2]+prediction[,4,3]*expectedyeild2014[4,2]
wheatland=prediction[,1,4]*expectedyeild2014[1,3]+prediction[,2,4]*expectedyeild2014[2,3]+p
        rediction[,3,4]*expectedyeild2014[3,3]+prediction[,4,4]*expectedyeild2014[4,3]
obx2014=array(dim=c(4,3))
for(m in 1:r){
 obx2014[m,1]=median(prediction[,m,2])
 obx2014[m,2]=median(prediction[,m,3])
 obx2014[m,3]=median(prediction[,m,4])
}

#setting up the various value that will change with each iteration price changes
predictedrevenue=array(dim=c(4,3,size))
lambdapredict=array(dim=c(r,size))
totalpredictedland2014=array(dim=c(r,size))
l2014changed=array(dim=c(r,size))
#cycle to determine the results of decreasing price
for(j in 1:duration){
for(m in 1:r){
predictedrevenue[m,,]=pricechanged[,j,]*expectedyeild2014[m,]
for(i in 1:size){
lambdapredict[m,i]=(predictedrevenue[m,,i]-cost2014[m,])%*%(obx[m,6,]/v2014[m])
if(m==1)slope=region1forcast$coef[2]
if(m==2)slope=region2forcast$coef[2]
if(m==3)slope=region3forcast$coef[2]
```

```
#slope for region four(Prairie Gateway) is set to zero due to it being negative
if(m==4)slope=0
totalpredictedland2014[m,i]=v2014[m]+slope*(lambdapredict[m,1]-datalambda2014test[m])
l2014changed[m,i]=(v[m,ceiling(t/2)]/totalpredictedland2014[m,i])^2
}


        for( i in 1:(size)){
        predictionchanged[i,m,,j]=modelfunctionfuturechanged(m,i,chaintotal[i,m,])
        }
}
cornlandnew=predictionchanged[,1,2,j]*expectedyeild2014[1,1]+predictionchanged[,2,2,j]*expe
ctedyeild2014[2,1]+predictionchanged[,3,2,j]*expectedyeild2014[3,1]+predictionchanged[,4,2,j]
*expectedyeild2014[4,1]
soylandnew=predictionchanged[,1,3,j]*expectedyeild2014[1,2]+predictionchanged[,2,3,j]*expec
tedyeild2014[2,2]+predictionchanged[,3,3,j]*expectedyeild2014[3,2]+predictionchanged[,4,3,j]*
expectedyeild2014[4,2]
wheatlandnew=predictionchanged[,1,4,j]*expectedyeild2014[1,3]+predictionchanged[,2,4,j]*exp
ectedyeild2014[2,3]+predictionchanged[,3,4,j]*expectedyeild2014[3,3]+predictionchanged[,4,4,j
]*expectedyeild2014[4,3]
#determine price due to demands reaction to total amount of crops avaiable
cornlandprice=(-(cornlandnew-cornland)/cornland/.91)*pricechanged[1,j,]
soylandprice=(-(soylandnew-soyland)/soyland/.38)*pricechanged[2,j,]
wheatlandprice=(-(wheatlandnew-wheatland)/wheatland/.75)*pricechanged[3,j,]
precentagechanged=matrix(c(cornlandprice,soylandprice,wheatlandprice),nrow=size,ncol=3)
pricechanged[,j+1,]=t(precentagechanged)+pricechanged[,j,]
cornland=cornlandnew
soyland=soylandnew
wheatland=wheatlandnew
}
name5=paste(name1,"predictionchangedown2")
name6=paste(name1,"pricechangesdown2")
save(predictionchanged, file=name5)
save(pricechanged, file=name6)

#redoing the same steps but now for a 20% increase in price
duration=25
prediction=predicttotal
predict=modelfunctionfuture(1,chaintotal[1,1,])

predictionchanged=array(dim=c(size,r,length(predict),duration))
pricechanged=array(dim=c(3,duration+1,size))
for(i in 1:size){
pricechanged[,1,i]=expectedprice2014*c(1.2,1,1)
}
```

```
cornland=prediction[,1,2]*expectedyeild2014[1,1]+prediction[,2,2]*expectedyeild2014[2,1]+pre
diction[,3,2]*expectedyeild2014[3,1]+prediction[,4,2]*expectedyeild2014[4,1]
soyland=prediction[,1,3]*expectedyeild2014[1,2]+prediction[,2,3]*expectedyeild2014[2,2]+pred
iction[,3,3]*expectedyeild2014[3,2]+prediction[,4,3]*expectedyeild2014[4,2]
wheatland=prediction[,1,4]*expectedyeild2014[1,3]+prediction[,2,4]*expectedyeild2014[2,3]+p
rediction[,3,4]*expectedyeild2014[3,3]+prediction[,4,4]*expectedyeild2014[4,3]
obx2014=array(dim=c(4,3))
for(m in 1:r){
 obx2014[m,1]=median(prediction[,m,2])
 obx2014[m,2]=median(prediction[,m,3])
 obx2014[m,3]=median(prediction[,m,4])
}

predictedrevenue=array(dim=c(4,3,size))
lambdapredict=array(dim=c(r,size))
totalpredictedland2014=array(dim=c(r,size))
l2014changed=array(dim=c(r,size))
for(j in 1:duration){
for(m in 1:r){
predictedrevenue[m,,]=pricechanged[,j,]*expectedyeild2014[m,]
for(i in 1:size){
lambdapredict[m,i]=(predictedrevenue[m,,i]-cost2014[m,])%*%(obx[m,6,]/v2014[m])
if(m==1)slope=region1forcast$coef[2]
if(m==2)slope=region2forcast$coef[2]
if(m==3)slope=region3forcast$coef[2]
if(m==4)slope=0
totalpredictedland2014[m,i]=v2014[m]+slope*(lambdapredict[m,1]-datalambda2014test[m])
l2014changed[m,i]=(v[m,ceiling(t/2)]/totalpredictedland2014[m,i])^2
}

        for( i in 1:(size)){
        predictionchanged[i,m,,j]=modelfunctionfuturechanged(m,i,chaintotal[i,m,])
        }
}
cornlandnew=predictionchanged[,1,2,j]*expectedyeild2014[1,1]+predictionchanged[,2,2,j]*expe
ctedyeild2014[2,1]+predictionchanged[,3,2,j]*expectedyeild2014[3,1]+predictionchanged[,4,2,j]
*expectedyeild2014[4,1]
soylandnew=predictionchanged[,1,3,j]*expectedyeild2014[1,2]+predictionchanged[,2,3,j]*expec
tedyeild2014[2,2]+predictionchanged[,3,3,j]*expectedyeild2014[3,2]+predictionchanged[,4,3,j]*
expectedyeild2014[4,2]
wheatlandnew=predictionchanged[,1,4,j]*expectedyeild2014[1,3]+predictionchanged[,2,4,j]*exp
ectedyeild2014[2,3]+predictionchanged[,3,4,j]*expectedyeild2014[3,3]+predictionchanged[,4,4,j
]*expectedyeild2014[4,3]

cornlandprice=(-(cornlandnew-cornland)/cornland/.91)*pricechanged[1,j,]
```

```
soylandprice=(-(soylandnew-soyland)/soyland/.38)*pricechanged[2,j,]
wheatlandprice=(-(wheatlandnew-wheatland)/wheatland/.75)*pricechanged[3,j,]
precentagechanged=matrix(c(cornlandprice,soylandprice,wheatlandprice),nrow=size,ncol=3)
pricechanged[,j+1,]=t(precentagechanged)+pricechanged[,j,]
cornland=cornlandnew
soyland=soylandnew
wheatland=wheatlandnew
}
name5=paste(name1,"predictionchangeup2")
name6=paste(name1,"pricechangesup2")
save(predictionchanged, file=name5)
save(pricechanged, file=name6)


#redoing the same steps but now for a 20% decrease in the yield for heartland

duration=25
prediction=predicttotal
predict=modelfunctionfuture(1,chaintotal[1,1,])
predictedrevenue=array(dim=c(r,crops))
pricechanged=array(dim=c(3,duration+1))
pricechanged=array(dim=c(3,duration+1,size))
for(i in 1:size){
pricechanged[,1,i]=expectedprice2014
}

cornland=prediction[,1,2]*expectedyeild2014[1,1]+prediction[,2,2]*expectedyeild2014[2,1]+pre
diction[,3,2]*expectedyeild2014[3,1]+prediction[,4,2]*expectedyeild2014[4,1]
soyland=prediction[,1,3]*expectedyeild2014[1,2]+prediction[,2,3]*expectedyeild2014[2,2]+pred
iction[,3,3]*expectedyeild2014[3,2]+prediction[,4,3]*expectedyeild2014[4,2]
wheatland=prediction[,1,4]*expectedyeild2014[1,3]+prediction[,2,4]*expectedyeild2014[2,3]+p
rediction[,3,4]*expectedyeild2014[3,3]+prediction[,4,4]*expectedyeild2014[4,3]

expectedyeild2014[1,]=expectedyeild2014[1,]*.8
obx2014=array(dim=c(4,3))
for(m in 1:r){
 obx2014[m,1]=median(prediction[,m,2])
 obx2014[m,2]=median(prediction[,m,3])
 obx2014[m,3]=median(prediction[,m,4])
}
predictedrevenue=array(dim=c(4,3,size))
lambdapredict=array(dim=c(r,size))
totalpredictedland2014=array(dim=c(r,size))
l2014changed=array(dim=c(r,size))
```

```
for(j in 1:duration){
for(m in 1:r){
predictedrevenue[m,,]=pricechanged[,j,]*expectedyeild2014[m,]
for(i in 1:size){
lambdapredict[m,i]=(predictedrevenue[m,,i]-cost2014[m,])%*%(obx[m,6,]/v2014[m])
if(m==1)slope=region1forcast$coef[2]
if(m==2)slope=region2forcast$coef[2]
if(m==3)slope=region3forcast$coef[2]
if(m==4)slope=0
totalpredictedland2014[m,i]=v2014[m]+slope*(lambdapredict[m,1]-datalambda2014[m])
l2014changed[m,i]=(v[m,ceiling(t/2)]/totalpredictedland2014[m,i])^2
}

        for( i in 1:(size)){
        predictionchanged[i,m,,j]=modelfunctionfuturechanged(m,i,chaintotal[i,m,])
        }
}
cornlandnew=predictionchanged[,1,2,j]*expectedyeild2014[1,1]+predictionchanged[,2,2,j]*expe
ctedyeild2014[2,1]+predictionchanged[,3,2,j]*expectedyeild2014[3,1]+predictionchanged[,4,2,j]
*expectedyeild2014[4,1]
soylandnew=predictionchanged[,1,3,j]*expectedyeild2014[1,2]+predictionchanged[,2,3,j]*expec
tedyeild2014[2,2]+predictionchanged[,3,3,j]*expectedyeild2014[3,2]+predictionchanged[,4,3,j]*
expectedyeild2014[4,2]
wheatlandnew=predictionchanged[,1,4,j]*expectedyeild2014[1,3]+predictionchanged[,2,4,j]*exp
ectedyeild2014[2,3]+predictionchanged[,3,4,j]*expectedyeild2014[3,3]+predictionchanged[,4,4,j
]*expectedyeild2014[4,3]

cornlandprice=(-(cornlandnew-cornland)/cornland/.91)*pricechanged[1,j,]
soylandprice=(-(soylandnew-soyland)/soyland/.38)*pricechanged[2,j,]
wheatlandprice=(-(wheatlandnew-wheatland)/wheatland/.75)*pricechanged[3,j,]
precentagechanged=matrix(c(cornlandprice,soylandprice,wheatlandprice),nrow=size,ncol=3)
pricechanged[,j+1,]=t(precentagechanged)+pricechanged[,j,]
cornland=cornlandnew
soyland=soylandnew
wheatland=wheatlandnew
}
name5=paste(name1,"predictionchangeyeild")
name6=paste(name1,"pricechangesyeild")
save(predictionchanged, file=name5)
save(pricechanged, file=name6)
```