2010

Three Essays in Environmental and Agricultural Issues

Sanchita Sengupta
Iowa State University

Three essays in environmental and agricultural issues

by

Sanchita Sengupta

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:
David A. Hennessy, Co-major Professor
Catherine L. Kling, Co-major Professor
Bruce A. Babcock
John A. Miranowski
Susana Goggi

Iowa State University

Ames, Iowa

2010

DEDICATION

I dedicate this dissertation to my husband Souma Chaudhury, my precious daughters, Rukma and Reina Chaudhury, my parents, Rita and Swapan Kumar Sengupta and my brother, Priyajit Sengupta. I love you all!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I express my sincere gratitude towards Dr. David Hennessy for his encouragement, patience and guidance. I have learned a great deal from him which will always be invaluable to me. His steadfast support, especially during the period when I was away from the university, always reinvigorated my enthusiasm and drove me towards the finish line. I also feel incredibly privileged to get the opportunity to work under the guidance of Dr. Catherine Kling. She has had a very strong influence on me and I admire her for her courage and enthusiasm. I am also very thankful to the rest of my committee members, Dr. Bruce Babcock, Dr. John Miranowski and Dr. Susana Goggi for their generous suggestions.

Rano Marupova has been a miracle of friendship and is undoubtedly the best thing to happen to me in Ames, Iowa. Mira Nurmakhanova, Ariun Ishdorj, Babatunde Abidoye and Majd Abdulla were wonderful and kind friends who sat through my presentations and helped shape my dissertation. I thank Sonam Gupta, Zeba Basu and Ira Payosova, my friends from Tucson, Arizona who were continuously encouraging me. A lively community of friends in Portland, Oregon where an invaluable support during the final days of this dissertation. Piu Banerjee and I sailed the same boat from India a decade ago. I am grateful for her intelligence and friendship.

I thank Rita and Swapan Kumar Sengupta for being wonderful parents in every possible way and I appreciate all the sacrifices they have done for me. Their patience, continuous support, inspiration and encouragement were an important anchor all these years. I am grateful to my amazing brother, Priyajit Sengupta, who is always ready to cheer me up and my aunt Mita Joglekar for her love and timely phone calls at key moments in my life. I also thank Karabi and Jiban Kumar Chaudhury, for their help and encouragement and for being wonderful grandparents.

It is hard to imagine how two little girls could contribute towards a doctoral thesis but they are right here with me, with their incredible cuteness, reminding me of all the great things that are yet to come. I thank Rukma Chaudhury for giving me the most wonderful experience of being her mother. And I thank my little daughter, Reina Chaudhury for being such a bundle of joy and laughter.

That this dissertation is finally complete is a proof of the endless encouragement, unconditional love and the never ending paper work and revisions delivered by my extraordinary husband. Thanks Souma, for being there for me and sharing this entire journey with me.

1. Avoiding biases from data-dependent specification search.

Abstract

The study evaluates the gains from avoiding data-dependent specification search on an estimation sample in an application to discrete choice models. We incorporate data splitting, the process by which the total available sample is randomly split in two or more sub-samples with the first (specification) sub-sample used for specification search, and the second (estimation) sub-sample used for obtaining "clean' estimates using the model chosen on the specification sub-sample according to a set criterion. We estimate 14 binary Logit models of the adoption of conservation tillage corresponding to the major sub-watersheds of the Upper Mississippi River Basin. For each of the sub-watershed models, we use the specification sub-sample to choose the explanatory variables that lead to the highest number of correct predictions provided that estimated coefficients are in conformity with economic theory. To evaluate the gains from avoiding specification search on the estimation sub-sample, we follow [33] and calculate the expected excess error, which is a measure of excess optimism concerning model fit on the specification sample. We find that the excess optimism varies with the sub-watersheds and has a tendency to be larger for the sub-watersheds with smaller samples.

Introduction

Estimation of econometric model parameters customarily assumes that the model structure is known. However, economic theory oftentimes provides only a partial guidance on the model structure, leaving the choice of the model's functional form and/or the set of explanatory variables to the researchers. This model uncertainty then leads to specification search by which explanatory variables are selected into the model to provide the best model specification according to preset criteria. However, if the same sample is used for both selecting the model and for fitting the model and making inferences, too narrow prediction intervals and biases in parameter estimates can ensue (Chatfield, 1995). In consequence, coefficient estimates and standard errors following pretesting cannot be used for valid inference ([96], [79]). Although the presence of non-trivial biases that result from data-dependent specification search is widely recognized by statisticians ([14], [61]), it is rarely taken into account in applied econometrics. Some exceptions to this practice are [18] and [49], who take into consideration the bias in inferences that arise due to specification search.

Admittedly, model uncertainty is difficult to quantify. The commonly proposed remedial approaches include the Bayesian Model Averaging Approach, collection of more data, and data splitting (see, e.g., [14]). This study focuses on data splitting, the process by which the total available sample is randomly split in two or more sub-samples with the first (specification) sub-sample used for specification search, and the second (estimation) sub-sample used for obtaining "clean" estimates using the model chosen on the specification sub-sample according to a set criterion. The other sub-samples (if any) are then used to further evaluate model fit. Since data sets available to researchers are almost never of the size permitting such procedure, this approach is rarely used in applied work and the studies reporting specification search biases are similarly scarce. Our analysis aims at filling this gap by evaluating the excess optimism concerning model fit attributable to data-specification search on the estimation sample in an application to discrete choice models.

In this paper we perform systematic data analysis and investigate the effects of data-dependent specification search for a data set that originally contains some 37,000 data points. We incorporate data splitting to estimate several binary logit models of the adoption of conservation tillage corresponding to major sub-watersheds of the Upper Mississippi River Basin, and estimate the excess optimism concerning model fit that is attributable to the data-specification search, using the approach developed by Gong (1986).

The rest of the paper is organized as follows. In section 2, we discuss why model uncertainty could be a problem and the different ways that have been used to deal with this problem.

Section 3 presents an empirical application to the estimation of discrete choice models of conservation tillage adoption, and section 4 concludes.

Model uncertainty

Pretesting or preliminary testing of the data to determine the type of model that is likely to be applicable, is a potential problem in statistics. Pre-testing could entail a coefficient restriction, testing for heteroscedasticity or serial correlation or as in our case, searching for the model with the largest number of correct predictions. [105] provide asymptotic results for inference after selecting a linear regression model based on final error prediction criterion. He finds the asymptotic variance to be satisfactory but asymptotic confidence regions to be too small. The problem is aggravated for small samples. But large sample with excessive data mining is also likely to lead to invalid inference. The Optimism Principle defined by [77], that model fitting necessarily gives optimistic results, is a manifestation of model uncertainty.

There are two schools of thought on the approach to dealing with model uncertainty, Bayesian and frequentist. Bayesian Model Averaging requires taking the weighted average of candidate models. The weights used are the Bayesian posterior probabilities and since they depend on the specification of prior probabilities, they are difficult to compute especially where there is no true model. Further, if the population form is uncertain, computing the Bayes factor could be another problem. We employ a frequentist approach in this study.

In the spirit of scientific inference which 'involves collecting many sets of data and establishing a relationship which generalizes to different conditions' (Chatfield, 1995), the ideal frequentist approach to solving model uncertainty is to use an existing data set for model selection through testing and then collect new data to estimate the selected model. However, collecting more data is expensive in most economic studies. A viable alternative to collection of new data to perform out-of-sample inference is data splitting.

Data splitting and model selection

According to [29], if a large data set is available, the best way to perform out-of-sample analysis is by a three-way random data split. The first set (specification set) should be used for selection of model, the second (estimation set) for estimation of the parameters and for point prediction and the third (validation set) for assessing the variability of the predictions. However, Faraway (1998) has noted that ‘the purpose of data splitting is to obtain better estimates of the variability of predictions, and the price one pays is that the actual variability of the predictions will tend to be higher’ as the size of the estimation sample is smaller than that of the original sample.

An important step in model selection is the selection of a criteria. There is no universally acceptable model selection criteria in the discrete choice models, but two common approaches are to select models with largest value of pseudo R^2 and the largest number of correct predictions ([97]). The goodness-of-fit statistic that is used in this study for specification search is the "percent correctly predicted". Specifically, we assume that a choice is correctly predicted if the predicted probability of the choice is greater or equal to 0.5. The threshold of 0.5 is not suitable for every discrete choice model (see, e.g., a discussion in [73]), but it works in our situation, since, as will be made clear from the application below, the cost of misclassifying one alternative is not very different from the cost of misclassifying the other alternative. In this paper, we first split the data set by applying the algorithm suggested by Faraway (1998) and choose the best fitting model based mostly on the goodness-of-fit criterion. We then use bootstrap methods to assess the benefits of avoiding specification search on the estimation sample.

Bootstrap methods for estimating excess optimism

To estimate the excess optimism concerning model fit that is attributable to data-dependent specification search, we employ bootstrap (resampling) techniques originally developed to correct for the optimism when data splitting is not an option ([23], [24] and [25]). As Efron and Gong (1983) point out, although theoretical basis for these methods is limited, the techniques can be successfully used in practice. The methods are based on the assumption that the original data set represents the underlying population and random draws from the original sample are draws from the same population.

The estimation of the excess optimism is based on the following observation (Efron,1982). Since the criteria for selecting the binary choice model with the best fit is the largest number of correct predictions, the prediction error or the apparent error is the number of incorrect predictions. Thus, the model selection bias can be manifested in the optimistic value of this apparent error. We follow Gong (1986) who proposed bootstrap methods to estimate the expected excess error.

Application

Agriculture in the Midwest has been targeted for conservation practices by various federal and state incentive-based programs. To better estimate the costs of current and intended programs and to better target conservation program expenditures there is an imperative need to understand the farm-level costs of conservation practices adoption for large, diverse areas. This study estimates these costs for one of the most effective conservation practices, conservation tillage (CT), for the entire Upper Mississippi River Basin (UMRB), an area which encompasses parts of Iowa, Illinois, Missouri, Wisconsin and Minnesota. The methodology we apply builds upon the work of [59] who estimate the costs of CT adoption for the state of Iowa.

Study region and data

The study region, the Upper Mississippi River basin (UMRB) is defined as U.S. Geological Survey hydrologic region 07 (http://water.usgs.gov). UMRB covers 492,000 square kilometers in parts of Iowa, Illinois, Missouri, Wisconsin and Minnesota. The entire basin is divided into sub-watersheds or 4-digit hydrologic units (HUC) that indicate the hydrologic region (first two digits) and hydrologic subregion (second two digits). There is substantial heterogeneity across the UMRB in terms of land use. As can be seen from Figure 1.1, the percentage area that is under cropland ranges from a minimum of 9.9% in HUC 7030 to 68% to HUC

7020. Incidentally, the major parts of both of these HUCs are in Minnesota. To reflect this heterogeneity, we estimate several CT adoption models corresponding to the sub-watersheds.

The data comes primarily from the Natural Resource Inventory (NRI) ([74]). The NRI is a scientifically based, longitudinal panel survey of soil, water, and related resources, designed to assess conditions and trends every five years. The 1997 NRI provides results that are nationally consistent for all nonfederal lands for four points in time 1982, 1987, 1992, and 1997. However, conservation tillage information is provided only in 1992 and hence only the 1992 data set is used for this study. The NRI data set for the UMRB region consists of a total of 103,849 observations. Table 1.1 shows the distribution of these points across the 4-digit HUCs and under corn, soybean production and conservation tillage. Most of the UMRB area is under corn production. Consistent with climate conditions, the northern HUCs have fewer soybean acres than the southern HUCs and tillage adoption is higher in the south than in the north. The NRI data set further provides information on geo-physical properties of the land, i.e. soil characteristics, slope, erodibility, and the like. The complete data set is formed by adding constructed net returns, climatic data and farm characteristics as in [59].

The economic theory provides a guidance only on which groups of variables ought to be present in the set of explanatory variables (such as the crop grown, soil and landscape characteristics of cropland, farmer characteristics, and climatic variables), and for the sake of brevity, we refer interested readers to [59] for the details on the rationale for each of the groups of the variables. Table 1.2 provides variable descriptions and summary statistics for the combined data set.

Adoption models

The models that are similar to that of [59] are derived under the assumption that a farmer adopts conservation tillage if the expected annual net returns from this farming practice, π_1, exceed those from the alternative, conventional tillage, π_0, plus a premium, P, associated with uncertainty. Then, assuming that $\pi_1 - P$ is a linear function of a set of observed explanatory variables x and that the observations on π_0 are available, the model is given by

$$\Pr[Y=1] = \Pr[\pi_1 \geq \pi_0 + P + \sigma\epsilon] = \Pr\left[\epsilon \leq \frac{\beta' x}{\sigma} - \frac{\pi_0}{\sigma}\right], \tag{1.1}$$

where ϵ is a logistic error and the observed dependent variable Y takes on the value of 1 of CT is adopted and zero otherwise. The parameters of interest are the linear function parameters β together with σ, the error term multiplier.

The specific models for each of the sub-watersheds are the variants of the basic specification, where

$$\begin{aligned}
\beta' x = {} & \beta_0 + \beta_{0,c} I_c + \beta_{0,s} I_s \\
& + \beta_1 SLOPE + \beta_2 PM + \beta_3 AWC \\
& + \beta_4 EI + \beta_5 OM + \beta_6 PH \\
& + \beta_7 TMAX + \beta_8 TMIN + \beta_9 PRECIP \\
& + \beta_{10} TENANT + \beta_{11} OFFARM + \beta_{12} AGE \\
& + \beta_{13} MALE + \beta_{14} CODE \\
& + PRSTD(\beta_{15} + \beta_{16}\pi_0 + \beta_{17} TENANT \\
& + \beta_{18} OFFARM + \beta_{19} AGE + \beta_{20} MALE \\
& + \beta_{21} CODE)
\end{aligned}$$

where, I_c and I_s are respectively the corn and soybean acres planted and are endogenous variables. The rest of the terms in the above equation are explained in Table 1.2. In addition to the specification described above, we also consider a specification that describes the probability of adopting conservation tillage as a function of the difference in the net returns between conventional and conservation tillage. In this case, instead of viewing the returns to conventional tillage as being known and that to conservation tillage being unknown, it is assumed that the average returns to both tillage methods are known. In this case, the model can be written as

$$\Pr[Y=1] = \Pr[\pi_1 \geq \pi_0 + P + \sigma\epsilon] = \Pr\left[\epsilon \leq \frac{\beta' x}{\sigma} - \frac{\pi_{0-1}}{\sigma}\right], \tag{1.2}$$

where π_{0-1} denotes the difference in net returns to conventional and conservation tillage. In this specification, $\beta' x$ represents the negative of the risk premium, rather than the difference between the expected net returns from conservation tillage and the risk premium. We refer to models (1.1) and (1.2) as net returns (NR) and difference (D) models, respectively.

Results of specification search

To conduct specification search, we split the sample of each HUC randomly into 4 sub-samples, and use the first sub-sample (specification sample), for specification search. In this search, we choose the specification that leads to the highest number of correct predictions, provided that the estimate of $1/\sigma$, which is the negative of the estimated coefficient of π_0 in the NR model and is the negative of the estimated coefficient of π_{0-1} in the D model, is positive as required by the theory. In this way, we find the best model structure and then obtain specification-search-bias-free estimates for the chosen models on the second (estimation) sub-sample. We chose the best-fitting models by varying the following model specifications:

1. *Area*: for each HUC, we choose the contiguous area containing the HUC,

2. *Variable*: choice among different soil and farmer characteristics variables,

3. *Model*: choice between the NR and D models.

Tables 1.3, 1.4, 1.5 and 1.6 provides parameter estimates and their standard errors after specification search. (on the estimation sample). Table 1.7 provides the percentages of correct predictions for the following four combinations of parameter estimates and data sets:

1. Specification sample and parameter

2. Estimation sample and parameter

3. Specification parameter and estimation sample

4. Estimation parameter and validation sample

Computing excess optimism

To estimate the excess optimism concerning model fit that is attributable to the data-specification search, we follow Gong (1986). Specifically, we consider the observed sample, $\mathbf{Z}_1 = (y_1, \mathbf{X}_1), ..., \mathbf{Z}_N = (y_N, \mathbf{X}_N)$ as being independent and identically distributed from an unknown distribution F. Here matrix $\mathbf{X}$ is defined as $\mathbf{X} = \begin{pmatrix} x \\ -\pi_0 \end{pmatrix}$ for the NR model, and as $\mathbf{X} = \begin{pmatrix} x \\ -\pi_{0-1} \end{pmatrix}$ for the D model. Let matrix β be defined as $\beta = \begin{pmatrix} \beta/\sigma \\ 1/\sigma \end{pmatrix}$. The prediction rule $\eta = \eta(\beta, \mathbf{X})$ associated with the model is the rule that allows predicting the value y_0 of the CT adoption indicator for any new set of observed explanatory variables $\mathbf{X}_0$. Let $e_0 = \beta'\mathbf{X}_0$. The prediction rule η is given by the following: $y_0 = 1$, if $exp(e_0) / (1 + exp(e_0)) > 0.5$, and $y_0 = 0$ otherwise.

Let us define $Q(y_0, \eta(\beta, \mathbf{X}_0))$ as the criterion that scores the discrepancy between the observed value y_0 and its predicted value $\eta = \eta(\beta, \mathbf{X}_0)$, which takes on the value of one if the observed and the predicted values are different, and zero otherwise. Let $\hat{F}$ be the empirical distribution function that puts mass $1/N$ at each point $\mathbf{Z}_1, ..., \mathbf{Z}_N$. The *true error* is defined to be the expected error that the set of estimates makes on a new observation $\mathbf{Z}_0 = (y_0, \mathbf{X}_0)$ from distribution F, $q = q\left(\hat{F}, F\right) = E_{z_0 \sim F} Q(y_0, \eta(\beta, \mathbf{X}_0))$. The *apparent error* of η is defined as $\hat{q}_{app} = q\left(\hat{F}, \hat{F}\right) = E_{z_0 \sim \hat{F}} Q(y_0, \eta(\beta, \mathbf{X}_0)) = \frac{1}{N} \sum_{i=1}^{N} Q(y_i, \eta(\beta, \mathbf{X}_i))$. Finally, the difference $R\left(\hat{F}, F\right) = q\left(\hat{F}, F\right) - q\left(\hat{F}, \hat{F}\right)$ is the *excess error*, and the expression $r = E_{\hat{F} \sim F} R\left(\hat{F}, F\right)$ is the *expected excess error* of the prediction rule $\eta = \eta(\beta, \mathbf{X})$. Here the expectation is taken over $\hat{F}$, which is obtained from $\mathbf{Z}_1, ..., \mathbf{Z}_N$ generated by F. If no data-dependent specification search has been conducted then the expected excess error is zero. However, if data-dependent specification search has been performed then the expected excess error is positive and thus is a reasonable measure of the excess optimism concerning model fit.

The bootstrapping procedure to compute the measure of optimism evolves in the following steps:

1. Let N be the number of observations in the sample $\mathbf{Z} = \{\mathbf{Z}_1, ..., \mathbf{Z}_N\}$. Take N random draws with replacement from $\mathbf{Z}$. These constitute one bootstrap sample, $\mathbf{Z}^b$. Estimate the selected logit model on the sample and obtain the bootstrap estimate $\hat{\beta_b}$.

2. Compute predicted probability with bootstrap estimates $\hat{\beta}_b$ and bootstrap sample explanatory variables $\mathbf{X}^b$ as $Y^*_{bi} = \frac{exp(\hat{\beta}_b \mathbf{X}^b_i)}{(1+exp(\hat{\beta}_b \mathbf{X}^b_i))}$ for $i = 1.....N$.

3. Compute predicted probability with bootstrap estimates $\hat{\beta}_b$ and the original sample $\mathbf{X}$ as $Y^*_{obi} = \frac{exp(\hat{\beta}_b \mathbf{X}_i)}{(1+exp(\hat{\beta}_b \mathbf{X}_i))}$ for $i = 1.....N$.

4. Apply the prediction rule η with the 0.5 threshold and obtain the proportion of incorrect predictions for both predicted probabilities, $q_{b0} = \frac{1}{N}\sum_{i=1}^{N} Q_{(bo)}$ and $q_b = \frac{1}{N}\sum_{i=1}^{N} Q_{(b)}$, where Q_{bo} is estimated using Y^*_{obi} and Q_b is estimated using Y^*_{bi}.

5. Repeat 1, 2, 3 and 4 a large number B times.

6. Obtain the estimate of the expected excess error, which is the average of the difference between two proportions taken over all bootstrap samples as $\omega = \frac{1}{B}\sum_{b=1}^{B}[q_{b0} - q_b]$.

Table 1.8 reports the estimates of the average error and the distribution of the measure of optimism ω over 1,000 bootstrap samples, for 3 different watersheds, HUC 7080, HUC 7100, and HUC 7110 with 1,641, 856, and 412 observations in the specification data set, respectively. Somewhat surprisingly, we get little difference in the model fit between the specification and estimation samples. An average error of 0.33 for HUC 7080 means that 33% of the time we get wrong predictions with the specification sample, while with estimation sample we get wrong prediction 32% of the time. If we correct for the optimism by adding the expected excess error estimates to the apparent error rates we get the bias corrected estimates as 34% for the specification sample and 33.5% for estimation sample.

Excess error results from computing the difference between the average number of incorrect predictions using the original sample and the bootstrap estimates, and the average number of incorrect predictions using the bootstrap samples and bootstrap estimates. The mean value of the optimism measure is positive, indicating that the apparent error tends to underestimate the prediction error. The magnitude of optimism is small, indicating that bias in the point estimate from data mining is probably not serious in our application, but it gets worse as the sample size gets smaller. The mean value is higher for the estimation sample than that of the specification sample. This shows that the specification search leads to better fit and hence a lower value of the optimism. Since the number of correct predictions is higher for specification sample than for the estimation sample, the number of incorrect predictions, conversely, should be lower for the specification sample resulting in lower values of the optimism parameter. Also, the values are consistent with increasing sample size. As the sample size becomes smaller the optimism parameter tends to be higher.

An Extension

The model presented in this paper could be used, for example, to compute regional-average subsidies that would provide estimates of the cost of adopting conservation tillage practices. Since we have four estimates from the four data combinations, it would be useful to evaluate which combination is most suitable for this purpose. This section proposes such an extension to the model.

The use of calibration techniques is a well known way to judge how good is a probability estimate. Calibration is a test of whether an assigned probability agrees with its relative frequency, *ex post*. The mean probability score or the Brier score is an alternative metric for evaluating probabilistic forecasts which compares the probability of an outcome with the actual outcome. One advantage of Brier score over calibration is that the Brier score can be decomposed into components that index both calibration and resolution, that is the ability of the forecaster to distinguish between events that occur and the events that do not occur.

Let Y be the actual binary outcome of the event. In the case of the tillage model, Y takes on the value of 1 if CT is adopted and zero otherwise. Y^* is the probabilistic prediction of the event. Then the quadratic probability score for a single observation or (forecast) is:

$$PS(p,d) = (Y - Y^*)^2 \tag{1.3}$$

PS ranges between 0 and 1. A score of 0 means perfect prediction, while a score of 1 is a bad prediction. This measure is different from the square of the correct predictions.

The mean probability score or Brier score ($\bar{PS}$) is an average of the single prediction version of the probability score over N occasions, indexed by $i = 1 \ldots N$:

$$\bar{PS}(Y^*, Y) = \frac{1}{N} \sum_{i=1}^{N} (Y_i - Y_i^*)^2 \tag{1.4}$$

Yates' Covariance Decomposition Calibration does not measure the ability of the forecaster to sort or distinguish between events that actually occur and events that do not occur. The Yates-partition of the Brier score is able to provide information on such sorting. Yates (1982) noted that the mean PS can be factored into its covariance decomposition:

$$\bar{PS}(Y^*, Y) = Bias^2 + Scatter + var(Y) + minvar(Y^*) - 2Cov(Y, Y^*) \tag{1.5}$$

where $minvar(Y^*)$ is the minimum forecast variance. In order to obtain the lowest $\bar{PS}$, the forecaster needs to minimize the square of the bias, Scatter, variance and minimum forecast variance terms and maximize $2Cov(Y, Y^*)$. All the terms in Equation 1.5 are explained below.

$Var(Y)$ represents the variance of the outcome index, defined as:

$$Var(Y) = \bar{Y}(1 - \bar{Y}) \tag{1.6}$$

where, $\bar{Y} = 1/N \sum_{i=1}^{N} Y_i$. $Var(Y)$ reflects the factors that are out of the forecaster's control. The remaining terms reflect factors that are under the forecaster's control.

$$\begin{aligned} Bias &= \bar{Y^*} - \bar{Y} \\ Cov(Y, Y^*) &= Slope * Var(Y) \\ Slope &= \bar{Y}_1^* - \bar{Y}_0^* \end{aligned}$$

where, $\bar{Y}_1^*$ is the conditional mean probability of adopting and $\bar{Y}_0^*$ is the conditional mean probability of not adopting.

$$\begin{aligned} Scatter(Y^*) &= \frac{1}{N}[N_1 Var(Y_1^*) + N_0 Var(Y_0^*)] \\ Var(Y_1^*) &= \frac{1}{N_1}\sum_{i=1}^{N_1}(Y_{i1} - Y_1^*)^2 \\ Var(Y_0^*) &= \frac{1}{N_0}\sum_{i=1}^{N_0}(Y_{i0} - Y_0^*)^2 \end{aligned}$$

Bias quantifies whether the probability predictions are too low or too high. It reflects the overall miscalibration of the forecast. $Bias^2$ reflects the calibration error regardless of the direction of the error. Scatter is interpreted as an index of general excess variability contained in the forecaster's judgements. The scatter statistic indexes the forecaster's responsiveness to information not related to event's occurrence.

The covariance measures the responsiveness of the forecaster to information related to the event's occurrence. The maximum value of slope is 1 which occurs when the forecaster always reports $Y_1 = 1$ whenever the event does occur and $Y_0 = 0$ whenever the event does not occur. The covariance term reflects the model's ability to make distinctions between individual occasions in which the event occurs or does not occur.

$Minvar(Y^*)$ is the minimum forecast variance defined as:

$$minvar(Y^*) = Var(Y^*) - Scatter(Y^*) \tag{1.7}$$

It represents the overall variance in the forecaster's probabilities if there were no scatter about the conditional means $\bar{Y}_1^*$ and $\bar{Y}_0^*$.

In the conservation tillage model, Y^* is the probability of adoption. The actual behavior is given by the variable $Till$

$$Y^* = \frac{exp(Estimate)}{1 + exp(Estimate)} \tag{1.8}$$

Table 1.9 reports the Brier score for HUC 7080 for each of the four combinations of parameter estimates and data sets. The Brier score for the estimation sample is minimum for specification sample since model uncertainty is least in this case. The specification sample estimation performs the best, as it is supposed to, mainly because of the high value of the covariance, reflecting the model's superior ability to make distinctions between individual occasions in which the event occurs or does not occur.

The out-of-sample validation performs marginally better amongst the remaining three estimation types, again mainly because of the covariance term. Bias is very low for all the estimation types, which indicates an overall good performance of the estimation. The variance of the actual outcomes Y or the exogenous factors affecting estimations remain more or less constant across the four estimations types. The scatter terms are highest for the specification and the out of sample estimation. The data set is common in these two cases, which probably explains the general variability in these two models.

The out-of-sample validation estimation performs well when presented under this criteria. Thus the subsidy estimates resulting from these out-of-sample validation would provide reasonable estimates as well as avoid the data-dependent specification search.

Conclusions

The objective of this paper is to evaluate the gains from avoiding data-dependent specification search on an estimation sample while estimating a number of conservation tillage adoption models for the Upper Mississippi river basin. We began by splitting randomly the total available data in four sub-samples. We undertook specification search on the specification sub-sample to select the models with the best fit. We then obtained the specification-search-bias-free estimates of model parameters by estimating the models selected on the second, estimation sample. Finally, we used bootstrapping techniques to estimate the measures of excess optimism concerning model fit. We found that the excess optimism is generally small, but varies with the sub-watersheds and has a tendency to be larger for the sub-watersheds with smaller samples. In the last section of the paper we provide a Brier score for the different combinations of the split data and the estimations. It is found that the out of sample validation estimation performs well in terms of minimizing different sources of error in the estimation process.

Because agricultural and ecological data sets are often characterized by a large number of observations, the model selection process we followed is viable for these data sets. While we did not find large gains from avoiding the improper specification search in our application, additional research is needed to evaluate the magnitudes of the gains in other applications. An interesting extension of this study would concern evaluating the gains of avoiding data-dependent specification search on the estimation of region-average subsidies needed for adoption of conservation tillage. As the estimates of the conservation tillage adoption model are affected by the specification search, so are the estimates of the subsidies which are functions of the data and the adoption model parameters.

Table 1.1 Description of the UMRB watershed by 4 digit HUC

4 Digit HUC	Total cropland points	Total area in million acres	Percentage of total area under cropland	Percentage of cropland area under corn	Percentage of cropland area under soybean	Percentage of total area under conservation till
7010	8954	1.2	18	61	4	2
7020	7797	0.92	69	50	28	12
7030	4113	0.46	10	67	1	2
7040	6495	0.65	33	69	6	14
7050	3847	0.55	11	70	1	4
7060	5930	0.55	42	78	6	32
7070	5141	0.66	14	66	1	5
7080	14965	1.46	67	62	24	45
7090	7167	0.66	56	78	9	22
7100	8375	0.9	64	54	28	43
7110	5883	0.59	44	35	19	14
7120	7661	0.63	55	58	22	18
7130	9745	1.13	72	57	29	26
7140	7776	0.79	44	42	19	13

Table 1.2 Descriptive Statistics

Notation	Description	Units	Mean	Standard deviation
Y	Conservation tillage(1-yes, 0-no)	Binary number	0.35	0.47
IC	Dummy variable for corn (1-corn,0-not corn)	Binary number	0.59	0.5
IS	Dummy variable for soybean (1-soybean ,0-not corn)	Binary number	0.34	0.47
π_{CVT}	Net returns to conservation tillage	$ per acre	88.88	81.62
SLOPE	Land slope	Percent	2.99	3.2
PM	Soil permeability	Inches per Hour	1.32	1.87
AWC	Available water capacity of soil	Percent	0.21	0.03
PH	Soil acidity (0 to 14) 7-neutral, less than 7 - acidic, greater than 7 - alkaline	Number	6.51	0.50
OM	Plant and animal residue in soil	Percentage	4.35	6.01
EI	Erodibility Index ($EI \geq 8$ are considered highly erodible land)	Number	5.75	9.78
TMAX	Mean of daily maximum temperature during growing season	Fahrenheit	78.5	2.82
TMIN	Mean of daily minimum temperature during growing season	Fahrenheit	55.4	2.95
PRECIP	Mean of daily precipitation during growing season	Inches	0.13	0.01
PRSTD	Standard deviation of precipitation	Inches	0.31	0.03
OFFARM	Proportion of operators working off-farm to the total number of farm operators	Number	0.5	0.06
TENANT	Proportion of harvested cropland operated by tenants to the total county harvested cropland	Number	0.18	0.07
AGE	County average farm operator age	Years	50.71	1.74
MALE	Proportion of male operators to the total number of farm operators	Number	0.97	0.01
CODE	Rural code for counties (0 to 9) 9 completely rural	Code	5.3	2.4

Table 1.3 Model specification and estimation

HUC	7010	7030	7050	7060	7070	7080
INTERCEPT	-4602.71	2643.45	1449.5	9845.19	-1344.09	3400.85
	(2092.38)	(835.1)	(596.26)	(6421.66)	(1644.82)	(1500.53)
CORN ID	15.33	5.2	10.38	21.04	33.64	6.32
	(10.68)	(3.60)	(4.15)	(17.14)	(15.54)	(5.72)
SOY ID	14.98	4.2	11.55	17.36	34.89	4.44
	(11.02)	(3.7)	(4.42)	(15.57)	(16.21)	(5.73)
SLOPE	-1.98	1.8	1.3	5.39	2.49	1.83
	(1.47)	(0.6)	(0.33)	(3.79)	(1.08)	(0.90)
PM	-1.33	-0.8	x	-2.41	x	-0.59
	(1.14)	(0.72)		(2.60)		(1.04)
AWC	7.25	-31.9	x	-192.07	x	-94.85
	(54.70)	(38.13)		(176.45)		(64.03)
EI	2.11	-0.32	x	-1.55	x	-0.31
	(1.34)	(0.2)		(1.13)		(0.28)
OM	-0.01	-0.07	x	0.32	x	0.11
	(0.28)	(0.16)		(0.56)		(0.23)
PH	-4.03	3.01	x	5.38	x	0.52
	(3.97)	(2.00)		(6.79)		(2.78)
TMAX	-5.39	0.14	x	10.55	x	0.25
	(2.76)	(0.6)		(7.11)		(0.94)
TMIN	6.20	2.23	x	-4.68	x	1.20
	(3.48)	(0.7)		(3.98)		(1.02)
PRECIP	-12.97	1118.9	1145.44	3134.21	2204.4	1243.24
	(401.89)	(228.6)	(230.74)	(1963.02)	(857.4)	(378.59)

x denotes variables that are not included in the estimation

Table 1.4 Model specification and estimation (continued)

HUC	7010	7030	7050	7060	7070	7080
TENANT	x	55.3	x	995.44	x	256.19
		(100.25)		(683.507)		(193.78)
OFFARM	x	52.9	x	-1049	x	59.81
		(105)		(831.25)		(230.74)
AGE	x	-3.6	x	-24.21	x	-1.55
		(3.7)		(17.51)		(5.25)
MALE	4740.74	-2896.5	-1649	-9539	1089.8	-3796.85
	(2145.77)	(827.2)	(62)	(6269.92)	(1632.9)	(1523.02)
CODE	x	8.6	x	13.90	x	14.44
		(2.7)		(11.83)		(5.35)
VPRECIP	-44780.9	28914.2	14013	105135	-14357.2	35780.70
	(20743)	(8446)	(5915.81)	(68027.8)	(16943.3)	(14916.50)
VRETURNS	-0.29	0.35	0.27	-0.063	0.8	-0.48
	(0.21)	(0.26)	(0.27)	(0.71)	(0.7)	(0.43)
VTENANT	x	297.4	x	9942.8	x	2537.93
		(1019.8)		(6726.62)		(1879.31)
VOFFARM	x	447.1	x	-9626.09	x	292.73
		(1040.5)		(7813.82)		(2087.97)
VAGE	x	-50.1	-14178	-292.74	x	-54.01
		(36.8)	(6066.1)	(206.35)		(51.77)
VMSHARE	45622	-27495	x	-90194.5	15262.9	-35052.1
	(21043.1)	(8143.8)		(59624.6)	(17411.1)	(14614.40)
VCODE	x	88.5	x	216.829	x	152.15
		(27.4)		(159.48)		(55.84)
Invsigma	14.68	13.7	16.42	43.80	36.29	17.38
	(6.93)	(2.6)	(3)	(28.83)	(14.38)	(5.47)

x denotes variables that are not included in the estimation

Table 1.5 Model specification and estimation (continued)

HUC	7090	7100	7110	7120	7130	7140
INTERCEPT	7742.55	8187.26	1932.89	483.212	1825.86	2851.59
	(7592.82)	(2573)	(4948.64)	(417.64)	(594.76)	(922.25)
CORN ID	86.96	0.72	4.58	13.94	17.21	20.26
	(85.88)	(6.46)	(8.46)	(8.05)	(7.87)	(6.96)
SOY ID	102.59	3.21	0.30	14.003	15.04	15.26
	(100.07)	(6.49)	(3.58)	(8.18)	(7.65)	(6.16)
SLOPE	8.40	0.40	1.15	3.60	3.85	2.27
	(8.05)	(0.57)	(2.10)	(0.87)	(1.06)	(0.71)
PM	x	-0.15	2.52	x	x	x
		(2.00)	(4.18)			
AWC	x	-76.05	142.96	x	x	x
		(68.74)	(241.51)			
EI	x	-0.05	-0.04	x	x	x
		(0.18)	(0.28)			
OM	x	-0.24	1.39	x	x	x
		(0.56)	(2.58)			
PH	x	-1.70	-1.67	x	x	x
		(1.77)	(3.67)			
TMAX	x	5.37	2.42	1.46	x	1.87
		(1.32)	(6.19)	(0.73)		(0.69)
TMIN	x	-7.13	-7.59	x	x	x
		(1.56)	(12.32)			
PRECIP	6430.86	630.07	-226.08	1305.36	1318.26	2285.15
	(6149.12)	(196.55)	(1060.02)	(286.47)	(331.28)	(593.49)

x denotes variables that are not included in the estimation

Table 1.6 Model specification and estimation (continued)

HUC	7090	7100	7110	7120	7130	7140
TENANT	x	569.60	-502.92	x	x	x
		(291.01)	(1220.52)			
OFFARM	x	347.68	-1229.81	x	x	x
		(221.71)	(1173.72)			
AGE	x	2.66	-36.86	x	x	x
		(11.35)	(64.87)			
MALE	-8804.91	-9023.52	1266.04	-791.50	-2048.74	-3320.78
	(8521.06)	(2762.26)	(2530.36)	(435.93)	(633.83)	(1004.06)
CODE	x	28.26	-22.71	x	x	x
		(9.75)	(18.62)			
VPRECIP	45345.2	81129.00	6045.21	3817.34	23497.9	36904.2
	(56301.8)	(25530.00)	(40530.3)	(4382.54)	(7203.35)	(10327.3)
VRETURNS	-2.87	-0.18	-9.52	0.13	-0.73	-1.09
	(7.34)	(0.26)	(1.04)	(0.22)	(0.22)	(0.29)
VTENANT	x	5006.36	-9276.06	x	x	x
		(2575.12)	(16185.7)			
VOFFARM	x	1802.43	-12706.9	x	x	x
		(2125.39)	(11795.7)			
VAGE	x	0.82	-316.16	x	x	x
		(96.73)	(599.85)			
VMSHARE	-45890.6	-86034.20	22212.6	-3554.98	-23554.6	-36694.6
	(57824.8)	(26886.6)	(30870.7)	(4488.29)	(7307.5)	(10410.9)
VCODE	x	245.67	-279.493	x	x	x
		(87.42)	(233.44)			
Invsigma	125.28	9.53	6.86	34.08	35.65	26.73
	(116.35)	(2.57)	(11.18)	(7.37)	(9.14)	(7.10)

x denotes variables that are not included in the estimation

Table 1.7 Goodness-of-fit measures

HUC	7010	7020	7030	7040	7050	7060	7070	7080	7090	7100	7110	7120	7130	7140
Area combinations that best fits the HUC	7010 7020 7030	7010 7020 7030	7030 7040 7080	7030 7040 7080	7040 7050 7080	7060 7080	7060 7070 7080	by itself	7080 7090	by itself	by itself	7080 7120 7130	7080 7130 7140	7080 7110 7130 7140
Model type	D	D	D	D	D	D	D	D	NR	D	NR	D	D	D
N	246	750	77	420	67	406	119	1641	680	856	412	660	1161	580
PCP with specification data	95.1	89.8	88.31	76.85	71.21	62.56	76.47	66.97	68.48	75.23	87.83	66.36	64.43	75.99
PCP with estimation data	95.53	87.15	89.61	75.47	50.74	66.25	84.03	67.64	65.73	74.65	83.49	66.51	63.74	73.45
out-of-sample PCP with specification and estimation data	93.49	87.15	93.5	75.47	68.65	61.33	84.87	61.24	66.17	72.31	85.19	64.7	63.74	73.79
out-of-sample PCP with estimation and validation data	95.95	86.24	85.9	74.52	69.11	61.67	81.6	69.36	66.47	75.38	80.34	66.87	61.53	71.55

PCP: percentage of correct predictions, D: Model specification with difference between the two returns
NR: Model specification with net returns from conventional tillage

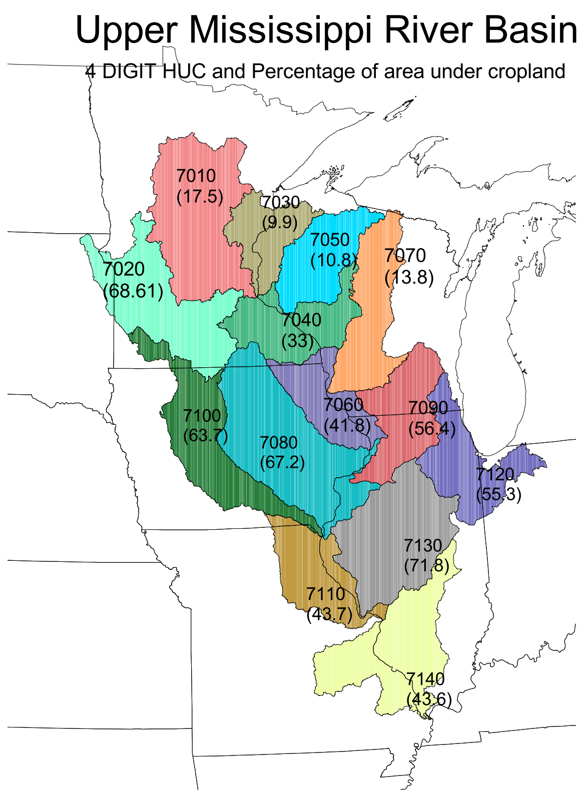
Table 1.8 Bootstrap estimation of the measure of optimism

Sample	Average error	Mean	Std. Dev.	Min	Max
Specification 7080	0.330	0.010	0.011	-0.028	0.05
Estimation 7080	0.323	0.012	0.011	-0.025	0.044
Specification 7100	0.248	0.015	0.013	-0.05	0.07
Estimation 7100	0.25	0.015	0.014	- 0.03	0.06
Specification 7110	0.12	0.019	0.016	-0.05	0.07
Estimation 7110	0.16	0.025	0.018	- 0.03	0.08

Table 1.9 Yates Decomposition of the Brier Score

Estimation Types	Brier Score	Bias Square	Variance of Till	Covariance	Scatter	Minimum variance of prediction
Specification	0.1975	0.000	0.2483	0.0513	0.0413	0.0106
Validation	0.2089	0.000	0.2479	0.0388	0.0326	0.0061
Out-of-sample	0.2207	0.00003	0.2479	0.039	0.0447	0.0061
Out-of-sample validation	0.2011	0.0002	0.2473	0.0435	0.0329	0.0076

Figure 1.1 4 digit Hydrologic Units in the Upper Mississippi River Basin



2. Empirical Ag-Environmental Model for Iowa

Abstract

This research develops an empirical model to estimate the effects of domestic policies like a) taxes on input prices, b)restrictions on input use and c)subsidies for adopting more environmentally benign crop rotation, on water quality and wildlife habitat. The crop acreage allocation, obtained from profit maximizing rotation decisions, affects the environmental variables. Results suggest that a subsidy payment for adopting a corn-soybean rotation has the largest impact on the improvement of water quality. Even though the policies become more expensive in the presence of rising commodity prices, the environmental pollution created by rising ethanol demand can be corrected with these policy instruments.

Introduction

Since the Uruguay Round Agreement Act (URAA) of 1994, which took effect in 1996, domestic agricultural policies have emerged as an important issue in multilateral trade negotiations. Current domestic subsidy policies in some countries may violate commitments made under URAA to reduce export subsidies, leading to negotiations at the WTO forum. An example of WTO influence stretching into what had formerly been viewed as a domestic policy is the A sugar quota in the European Union. The Dispute Settlement Body of WTO ruled in 2004 that price discrimination along with production quotas have the effect of cross subsidizing exports, ([20]). The A sugar quota would almost certainly be filled and the subsequent higher domestic price for sugar could then be viewed as a pure income transfer to farmers, leading to production beyond the quota amount at lower world prices. Decisions at the margin are also affected because some farmers limit their production to the quota amount and would have exited the industry were it not for the high domestic prices resulting from production within the quota. [15] examine the effect of infra-marginal production subsidies like loan deficiency payments (LDPs) or production flexibility contract (PFC) payments on exit and output expansion decisions beyond quota. They calibrate the production and cost structure of the United States wheat sector and find a large impact from removal of these payments on the exit decisions of the low profit farm units. But the aggregate impact on output is small so long as the marginal farm remains small.

Currently the WTO recognizes environmental protection as a legitimate policy goal. Domestic environmental policies are placed under the 'green box' which permits direct payments that are not linked to the present or future production of any specific product and is determined, instead, by historic levels of production. Green box programs are considered to be non-trade distorting and give complete freedom to the governments to provide unlimited amounts of funds under these programs. Conservation programs are an example of green box policies. The legality of these green box payments has been questioned during the 2003 dispute between US and Brazil over cotton subsidies ([89]). The case against the United States was that several of its domestic programs, that are included in the green box, are export subsidies that have depressed world cotton prices and increased US exports. Some countries are now proposing to place limitations on the amount of payments that can be provided under the 'green box' policies. Payments at issue include EQIP (Environmental Quality Incentive Program), which is a cost share program, and CRP (Conservation Reserve Program), which is a rental and easement payment program. These, rather costly programs, actually take land out of production which could be a reason why they still remain unchallenged in the WTO forum.

As with cotton and sugar disputes, the United States may face challenges with respect to some of its domestic environmental policies. [51] provides various reasons for these challenges to 'green box' policies. Firstly, it is not very easy to define what is 'minimally trade distorting' without extensive research. Secondly, any income support that reduces downside risks in fluctuations distorts production. Finally, any expectation of reassessment of the historical bases that determine payments can prevent farmers from adopting more environment friendly land uses. [9] provide an excellent literature survey on decoupling of farm programs. The case against US domestic support would gain more importance in the WTO forum if environmental policies are increasingly used to provide income support to farmers. Evidence on any improvements in environmental outcomes, linked directly to these policies could help to inform the debate. A production possibility frontier that incorporates all the environmental outcomes along with production and acreage allocation outcomes would give a complete picture of these agricultural-environmental linkages. Some control interventions in output and input markets may or may not be defensible from a pollution management perspective. Specifically, government support could be severely restricted if there is no environmental improvement in the presence of trade distorting issues. The question then is how well prepared are the parties involved in understanding the linkages between domestic support policies and the environment.

Establishing the linkages between market responses and changes in environmental variables becomes critical in the light of the new farm bill and forthcoming WTO negotiations. The [75] recommends that *"The combination of agricultural and environmental policy measures .. be carefully designed and implemented to ensure coherence so that they improve environmental quality in the most cost-effective and transparent way, with least distortion to production and trade."* In light of these recommendations, [55] examine the cost-effectiveness of 'green payment' policies for agricultural production. Green payments are financial incentives given to the farmers for voluntarily changing their choice of input use or production technology so that there is less pollution from agricultural production. [55] find that the costs of abatement of alternative green payment policies are not very different relative to each other, unless it is a large cost sharing program. Looking at other standards and pricing policy approaches, it is then necessary to base these policy choices by comparing their effects on environmental outcomes.

Another important issue concerning U.S. agriculture bio-fuel production. Ethanol usage has a large advantage in reducing green house gas emissions but might come in conflict with other environmental attributes. The [8] projections in 2006 show an increase of 6 million acres of the U.S. corn acreage during the next 10 years, attributed mostly to rising ethanol demand.

These changes will occur mainly in the Upper Mississippi River Basin (UMRB) and will have dramatic implications for water quality via nitrogen use and soil erosion, especially on the hypoxia zone of the Gulf of Mexico ([80], [30]). [83] in their popular Science magazine article claim that world-wide land use change, as a result of US policy, will nearly double greenhouse emissions over 30 years and increase greenhouse gases for 167 years.

As long as the United States uses corn as an input in the production of ethanol, corn demand and prices will increase. One way to meet this additional demand would be to reduce US corn exports to other countries. If this increased demand is met by more intensive corn production per acre then the increase in the allocation of corn acreage could counter the policies that provide adequate environmental benefits. Some bio-fuel promotion policies, for example, the ethanol excise tax exemption and the tax credit to retailers, are already effective in Iowa. These results in high demands for corn which, in turn, leads to high price commodity market scenario. It is crucial to take into account this market situation while analyzing policy effectiveness.

Several studies have looked at the impact of non-point agricultural pollution, specifically how crop acreage allocations affect environmental amenities. (e.g.[101],[99], [102])). These studies have incorporated economic models for estimating the crop choices, tillage choices, crop rotation choices and adoption of best management practices. The above studies, particularly the ones with crop rotation choices, do not take into account the yield effects of rotation on subsequent crops, mainly because the data set is simply not available to conduct such research. The decision to adopt a particular rotation at a particular time depends crucially on the profit maximizing yield and fertilizer requirements of all the rotations under the prevailing market conditions. Many agronomic studies have consistently shown the presence of a yield drag in a subsequent crop where corn after corn average yield is 10% less than corn after soybean average yield. These yield effects should be incorporated when estimating rotation adoption models.

Policies differ in their effect on the entry and exit decisions in the presence of heterogenous land quality. Several studies have addressed the effect on acreage/planting decisions in the presence of land set-aside and cost share programs. [100] has attempted to quantify the amount of non-cropland brought into crop production as a result of the conservation easements on the farmland. This slippage effect could be caused by increased output prices and substitution effects. [88] models the interaction between conservation and production decisions and shows how this interdependence causes unintended effects for a cost share program by inducing changes in land profitability and affecting cropping patterns. [55] emphasizes the role of information acquisition in the presence of heterogenous entities. The present paper overcomes the requirement to model entry and exit decisions by using aggregate panel data of county level land use. At this level of analysis the individual entry-exit decisions are integrated into aggregate acreage allocation.

The environmental literature has looked at multiple benefits of a single policy ([3], [60], [106]). Specifically, most studies have looked at the multiple benefits of more costly programs like the Conservation Reserve Program and Conservation Security Program. Although policies like fertilizer taxes and quotas attempt to target the immediate environmental concern of water quality benefits, they might have some residual effects due to changes in acreage allocation. Many studies with integrated models make use of physical models to simulate the environmental variables under different management conditions. An example of such a tool is the Soil and Water Assessment Tool (SWAT) which is being increasingly used to simulate environmental outputs. The inputs for these models, usually land topography, weather conditions and soil conditions, are many and remain constant throughout simulations. While this tool is desirable when monitoring data is not available and is efficient in addressing the complex network of rivers draining a watershed, it is limited in the scope of the environmental variables that can be addressed. The observations on other environmental variables like wildlife and lake water quality, will not be obtained with SWAT models.

The purpose of this paper is to develop a framework which helps in understanding the consequences of various alternative domestic policies for environmental management. The model has two components. First a framework for an ecological model that links acreage allocation to environmental variables is established to obtain an effective mapping of economic choices to environmental outcomes. The environmental outcomes that we propose to study are river and lake water quality and wildlife habitat. For certain time periods there are extensive monitoring data available on water and wildlife in Iowa. These environmental outputs can be incorporated into the production possibility frontier along with the shares of acres allocated to corn and soybean, the two main crops grown in Iowa. Factors like soil quality, weather, and other ecological variables are accounted for in estimating the effect of acreage allocation on environmental outcomes.

Acreage allocation is then determined by a model developed to study the market determinants of rotation choices. A major contribution of this paper is the way in which the economic model has been formulated. The theoretical foundation for this paper is given in [42] where the author emphasizes that the spill over yield effects persisting for one or more years, is important in determining the optimality of a particular rotation. Access to an agronomy field data set allows the estimation of the production technology for different rotations [1]. The estimated production technology is applied to each county after appropriate scaling. Under joint price and yield uncertainty, the profitability of various rotations are estimated for each county. Economic theory suggests and statistical analysis confirms that the revenues of corn and soybean yields are not independent. The correlation between them needs to be taken into account while generating the random deviates for profit estimations. An emerging tool in finance that is used in this analysis to address the correlation structure is the copula approach. Copulas are functions that join multivariate distribution functions to their one dimensional marginal distribution functions ([70]). We then proceed to find the probability of occurrence of the two most optimal rotations given various combinations of input and output prices in each county. The acreages allocated to corn and soybean production are then determined for all the counties. The producers are assumed to adopt the two most profitable rotations, at an aggregate level. This links the market prices and input use to the acreage allocation. The main advantage of this approach is that it includes effect of changes in yields in the estimation

[1]Data have been kindly provided by Antonio Mallarino, Department of Agronomy, Iowa State University and Ken Pacionovsky, farm superintendent. These are field trial data from Iowa State University's Northeast Iowa Research and Demonstration Farm located in Floyd County, IA.

of crop acreage under different rotations.

The above framework can then be used for evaluating different policy scenarios in improving environmental degradation. Policies addressed here are fertilizer taxes, a per acre subsidy for a corn-soybean rotation adoption and a restriction on nitrogen use. The estimates for expected nitrogen use, average profits, acreage allocation, and effects on water quality and pheasant population are provided. With baseline values of 2001, all the policies are compared under two price scenarios. These are low corn prices and high corn prices, in anticipation of the boom in ethanol market. The study provides important insight into policy effectiveness with respect to environmental output.

Ecological models

In Iowa, agricultural runoff has been identified as a primary source of water quality problems. Fertilizers are the leading cause of increases in nutrient levels in water bodies, beyond sustainable levels that are consistent with the ecosystem. Nitrate nitrogen in excessive amounts may cause lake eutrophication, depleting the level of dissolved oxygen necessary for sustaining aquatic life. River water quality in terms of nutrients in this region in the Upper Mississippi River Basin has direct impact on the hypoxia zone in the Gulf of Mexico ([80], [30] and [57]). Assessment of environmental outcomes as a result of a policy change requires accounting for all environmental outcomes, which is beyond the scope of this study. For example, we do not consider environmental benefits from carbon sequestration. However wildlife changes are also considered along with water quality changes in lakes and rivers. It will be shown later in this paper that the same policies have different effects on water quality and wildlife habitat, not only in their magnitude but also in the direction of change.

Lake water quality

Lakes provide environmental amenities for recreational use and drinking water supply. The key factors for promoting recreational use of lakes are improving water clarity, reducing algal blooms and increasing fish population. There is no unique water quality indicator that adequately encompasses all the requirements for different uses. The nutrient levels for lakes are appropriately determined by its assigned beneficial use. This paper looks at the levels of key nutrients in all the lakes regardless of their uses.

Extensive data on the lake water quality in Iowa are provided by the Limnology Laboratory at Iowa State University. According to the Report on Iowa Lakes Classification for Restoration, May 2005, the Iowa Lakes Survey was conducted by sampling 132 principal recreational lakes from 2000 to 2005. One hundred and fifteen of the lakes were previously studied, and classified for restoration, in 1979 and again between 1990 and 1992 ([6]). The Iowa Lakes Survey refined the existing classification system based on water quality, public benefit and potential restoration effectiveness.

Some of the water quality measures considered are nutrient levels like total nitrogen, total phosphorus and water clarity measures like total suspended solids (TSS) and Secchi disk depth, all of which are affected by agricultural pollution. Nitrogen from cropped fields is a leading source of water quality pollution in agricultural areas. Phosphorous is the key nutrient affecting the amount of algae growth. [1] find that intensive row crop agriculture leads to higher levels of N:P. Measurements for total phosphorus include soluble phosphorus and the phosphorus in plant and animal fragments suspended in lake water. Secchi depth is the depth at which the bottom of the lake can still be seen, providing a measure of water transparency. TSS provides the actual weight of particulate matter present in a sample of water collected from the site. Table 2.1 provides the summary statistics for the lake variables. The EPA's recommended nutrient criteria for lakes and reservoirs in the Corn Belt and Northern Great Plains Region are 37.5 μg/l of total phosphorus, 0.78 mg/l of total nitrogen and Secchi Depth of 1.36 m. The average lake water quality in Iowa is poorer than the EPA recommended criteria. The table indicates considerable variation across the panel data set, in terms of lake quality. The levels of lake nitrogen and phosphorus in 2001 are shown in Figures 2.1 and 2.2. Nitrogen levels are highest in north-central Iowa region which is the prime agricultural area. Phosphorus levels in lakes are more spread out without any particular pattern.

The following set of equations is estimated as the supply function of the lake water quality:

$$TotalNitrogen = a_0 + a_1 * corn + a_2 * soybean + b * x + c * z \quad (2.1)$$

$$TotalPhosphorus = a_0 + a_1 * corn + a_2 * soybean + b * x + c * z \quad (2.2)$$

$$TSS = a_0 + a_1 * corn + a_2 * soybean + b * x + c * z \quad (2.3)$$

$$SecchiDepth = a_0 + a_1 * corn + a_2 * soybean + b * x + c * z \quad (2.4)$$

where *corn* and *soybean* are the total acres of corn and soybean respectively, in the county. Corn acres, soybean acres, and the total land in farms were obtained from the National Agricultural Statistical Services (NASS) data for 99 counties in Iowa for the years 2000 to 2005. Vector x comprises lake characteristics like volume of the lake in cubic meters, the size of the lake watershed (in acres) and the average temperature of the lake. The lake characteristics were also obtained from the Iowa Lakes Survey. All of these factors are important in explaining the capacity to process the nutrient supply. The z vector comprises county characteristics like average precipitation, texture of soil (coarse or fine), percentage of land in the county that has maximum slope greater than 8%, weighted average of corn suitability rating (CSR)[2] for the county and the average value of K factor[3] for the county. These county characteristics are obtained from three sources, the Natural Resource Inventory (NRI, 1997), the Iowa Soil Properties and Interpretation Database (ISPAID, 2004) and the Iowa Environmental Mesonet (IEM) for the climate data . The lake water quality is mainly affected by runoff from the lake watershed which are distinct from the river watersheds and are much smaller. However, for simplicity it is assumed that the land use in the entire county affects the water quality at the lakes[4].

[2]CSR is an index that provides relative ranking of all soils based on their potential to be utilized for intensive row crop production. Therefore higher values of this variable is likely to contribute towards more polluted lakes simply because more row crop production is likely to take place in these lands.

[3]Higher values of K factor indicates higher susceptibility to rill and sheet erosion by water. The estimates are based primarily on percentage of silt, sand, and organic matter (up to 4%) and on soil structure and permeability. High crop production on land with high K Factor is expected to lead to higher lake pollution.

[4]Ideally a panel data set on land use in the lake watersheds should be used in this study. However, such a data set is not available. Instead, the panel data of land use in the counties is assumed to provide land use changes in the watershed. This serves the additional purpose of estimating the effect of county crop acreage on surface water quality.

River and stream water quality

A set of equations similar to lake water quality can be estimated for rivers and streams. One complication that might arise in using river and stream water quality data is the overlaying of watershed boundaries with county boundaries. Water quality attributes in rivers and streams are affected by the land use pattern in the watershed. Since this study is done at the county level, there is a need to address this disparity. Iowa cropland drains into either the Mississippi River in the east or the Missouri River in the west. Any river or stream that drain into either one of these rivers, is affected by land-use in the watershed defined by that stream. Although it is desirable to include only the upstream acres for a particular monitoring point, the data on the annual acreage allocation for smaller watersheds are not readily available. Instead, we include the set of areas defined by the intersection of the counties and the larger watershed that contains the monitoring site. The area under crop c that affects a site k, at time t is then given by:

$$y^c_{kt} = \sum_{i \in \Omega_k} W_i C^c_{i,t} \tag{2.5}$$

where, Ω_k is the set of areas formed by the intersection of county boundaries and the watershed boundaries that affect site k. W_i are the weights constructed by taking the area of each of the intersection as a percentage of watershed area and $\sum W_i = 1$. $C^c_{i,t}$ is the total area of the county under the particular crop at time t.

River and stream water qualities are obtained from monitored data maintained by Environmental Protection Agency (EPA). These data are are available at the web site of Iowa Department of Natural Resources. The Iowa Ambient Water Monitoring Program is a statewide monitoring program for Iowa's surface, groundwater, lake, and wetland resources. It is administered by the Geological Survey Bureau of the Iowa Department of Natural Resources. Table 2.1 provides summary statistics for the river water quality. The EPA's recommended nutrient criteria for rivers and streams in the Corn Belt and Northern Great Plains Region are 76.25 μg/l of total phosphorus, and 2.18 mg/l of total nitrogen. The average amount of nutrients in the rivers in Iowa exceeds these criteria.

The following equation is estimated for the river and stream water quality:

$$Nitrogen = a_0 + a_1 * y^{corn} + a_2 * y^{soybean} + b * flow + c * z \tag{2.6}$$

$$Phosphorus = a_0 + a_1 * y^{corn} + a_2 * y^{soybean} + b * flow + c * z \tag{2.7}$$

where y^{corn} and $y^{soybean}$ are the constructed total area of land in the watershed under corn and soybean respectively. Vector z comprises watershed characteristics that would affect river and stream qualities. Some of the watershed characteristics that are considered here are average slope of the watershed, the water erodibility index (EIwater) and wind erodibility index (EIwind). The watershed characteristics are obtained from the 1997 National Resource Inventory (NRI) data set.

There are 54 water monitoring sites used in the study. Nitrogen and Phosphorus levels are available in most cases as monthly observations for 8 years, 1999-2006. The dependent variables are the annual average of the observations. The explanatory variable, $flow$ is the amount of water flowing into the water body at the monitoring site. Figures 2.3 and 2.4 show the nitrogen and phosphorus levels at the selected water monitoring sites in the year 2001.

Wildlife habitat

An increase in pheasant population is viewed as an indicator of wildlife benefits. Hayfields, oat fields, pastures, idle grassland areas, wetlands and Conservation Reserve Program lands provide good pheasant nesting cover while intensive row cropping and habitat fragmentation is detrimental to pheasant population. [17] recognized grassland cover as one of the most important determinants of pheasant population and studied the effects of grassland under CRP as opposed to conversion of grassland to conservation buffer strips, on pheasant population. They showed the importance of a diversified agricultural landscape with large blocks of undisturbed habitat as opposed to disturbed habitat fragments to pheasant population.

The pheasant data are obtained from Iowa Department of Natural Resources (IDNR), which conducts an annual roadside survey of upland game population. The total number of pheasant counts on 30 mile routes in each county are provided from 1962 to 2005. The statewide pheasant counts were low for the years with severe winter or abnormally wet weather. [68], studies the impact of land use changes on the pheasant population in Iowa. A distinction is made between the northern row crop region and the southern pasture region, providing different wildlife habitats. As seen in Figure 2.5, the north-west region, which is the prime agricultural region in Iowa has most of the pheasants.

Almost 89% of the annual diet of ring-necked pheasant consists of seed, primarily corn even though corn and soybean are considered to be of little value to pheasant population as foraging habitat because of their low insect abundance and biomass. According to the Department of Conservation, Missouri, ring-necked pheasants do best where there are agricultural crops along with some grassland for nesting and woodland for winter cover. Also, pheasant food and cover are more diverse on farms using a crop rotation system.

The following supply equation for the pheasant population is estimated:

$$Pheasant\ Population = p_0 + p_1 * cornsh + p_2 * soybeansh + p_2 * cornsh^2 + p * z \qquad (2.8)$$

In the above equation, the share of agricultural land in corn, *cornsh* and its square are assumed to be explanatory variables in the determination of pheasant supply. The quadratic function was hypothesized by Clark et. al. (2002), since there seems to be an optimal amount of corn land that is most suitable for pheasant growth. Vector z in this case comprises of county characteristics like county population from the U.S. Census 2000, precipitation and a dummy variable for the counties located in the northwestern part of Iowa. County population is expected to have an indirect effect on pheasant population. Larger population might lead to more hunting of pheasants. Precipitation as a measure of rain is an important factor in pheasant habitat. Early rain is good for nesting cover while too much rain might limit the hatch. Finally the dummy variable distinguishing northwestern counties from the rest is included to take into account the natural and more favorable habitat in this area [68].

Ecological model estimation and discussion

Lake water quality

Table 2.2 provides the results for pooled SUR estimation of the lake water quality production function. The system weighted R^2 for this model is 0.2879. Most of the coefficients for Sechhi Depth are significant. Corn and soybean acres are not significant for any of the nutrients. Tests for multi-collinearity shows that the variance inflation factor for corn and soybean acres are a high of 11.64 and 16 respectively. This suggests that the corn and soybean acres might be correlated. Further tests indicate that for the years 2000-2005 the corn acres and the soybean acres are negatively correlated within a county over time, although the correlation is not significant for most counties. Table 2.3 presents the results for the SUR estimation without soybean acres. The R^2 for this model decreases marginally to 0.2857. We find in this table that corn area is a significant variable in the determination of lake quality. However, for our policy analysis, the estimation results of Table 2.2 are considered, since some of the policies shift land from corn to soybean production.

In both the specifications, most of the explanatory variables are significant. As expected, larger area of the lake watersheds, precipitation, sandy soil and soil phosphorus content are all significant pollution enhancing factors. Higher sloped lands and lands with higher corn suitability rating tend to decrease nutrients and clarity. This could be because there is less cultivation in highly sloped lands, whereas higher CSR indicates better quality land and hence less pollution.

River and stream water quality

Table 2.4 shows the estimations for the river water quality supply in Iowa. The R^2 for this system of equation is 0.0562. The area under corn and soybean production are significant variables in the determination of total nitrogen at the monitoring sites. However, these are not significant determinants of phosphorus levels. Nitrogen and phosphorus levels increase with corn area and the flow into the water body, and decrease with the soybean area. The flow into the rivers at the monitoring sites is defined as the portion of precipitation on the surrounding land that ultimately reaches streams, often with dissolved or suspended material. A positive coefficient on this variable is expected since the surrounding area is mostly agricultural. Estimations without the soybean acreage are provided in Table 2.5. Omitting the soybean acres from the estimation results in a positive and significant coefficients for corn acreage for nitrogen. Again, as in the case of the lake water quality, the estimation results with soybean acres are considered for policy analysis.

Pheasant population

The estimated coefficients of the pheasant supply function are given in Table 2.6. The dependent variable in the ordinary least squares regression is the average annual pheasant population per 30 mile route per county and the R^2 for this model is 0.071. The share of corn acres has a significant positive effect and the square of the shares of corn acres has a negative effect on the pheasant population. The average pheasant population reaches a maximum level at 38.75% share of corn acres in a county on an average, assuming all other variables remain constant. In the study by [17], the average pheasant population reaches a maximum of 54% of grassland or buffer strip, while the remaining 46% comprises of row cropland (primarily corn and soybean), roads, wetlands, hayland, pastureland and oats. Their study is based on two northwestern counties in Iowa. Palo Alto County has 57% and performs better than Kossuth county with 86% row crop, in terms of supporting pheasant population. The percentage acres of soybean in the county seems to have a negative effect on pheasant population. Population density has a significant negative effect on the pheasant population. The northwestern counties comprising of prime agricultural land are good for pheasant population as shown by [68].

Economic model

Most studies in the literature that perform environmental policy assessment analyze the adoption decision of specific management practices by farmers. For example [59] study conservation tillage adoption and [37] study the adoption of irrigation technology. These studies typically require data on the adoption decisions at the farm level, namely crop rotation decision of each farmer in the entire state. In the absence of such data set, an alternative procedure is employed. The results of a field experiment conducted in one county are extrapolated to all the other counties. The design of the field experiment allows the estimation of a primal production function incorporating the yield effects of input use. Together with market prices prevailing in the corresponding years, conditions for profit maximization can be estimated. A disadvantage of this method, common to all experimental data, is that actual farmers' decisions are not used. Nevertheless, this procedure contributes to the literature by addressing the joint effects of nitrogen input use and rotation on yield. The steps involved in the procedure are outlined briefly in the following paragraph.

First, crop distribution functions are estimated using field level agronomy data for crop yields under different rotations. A county-wide shift parameter is included to estimate rotation-specific yield distribution functions for each county. This takes into account the spatial heterogeneity of crop production across the counties. Second, the profits under different rotations are estimated for each county. A crucial assumption in this step is that the distribution functions (with the exception of the parallel shifts) that are estimated from field data hold true for all counties in Iowa. Under joint uncertainty in prices and yields, Monte Carlo methods are used to estimate average profits for each county under each rotation. Difference in the prices of corn and soybean across county are taken into account. The correlation between crop yields and prices are accounted for in the construction of random deviates, using a Copula approach. Finally, the probability of occurrence of the two most profitable rotations are computed for various combinations of input and output prices for each county. This provides a smoothing factor for the calculation of total land under corn and soybean production for each county. Use of this factor accounts for heterogeneity among producers and prevents all from switching to the single most profitable rotation upon a small change in market factors. The acreage allocated to corn and soybean production is then determined for all the counties. In the next three sections these steps are explained in detail.

Estimation of crop yield and price distribution function

The yield data under different rotations allow us to estimate density functions for generating observations for simulation. The data set is obtained from experiments on Iowa State University's Northeast Iowa Research and Demonstration farm located in Floyd County, IA. Twenty-five years of field trials were conducted since 1979, with 3 plot-level observations on yield per year. Five rotations were considered per field: Continuous corn (C), Continuous soybean (S), corn-soybean (CS), corn-corn-soybean (CCS) and corn-corn-corn-soybean (CCCS). For each of these rotations, crop yields were recorded at four different levels of nitrogen 0, 80, 160 and 240 lbs per acre. For estimation purposes, corn at different stages of rotation are taken to be separate crops. For example, for a CCCS rotation there are four crops, first year corn, second year corn, third year corn and soybean and four different density functions for each of these crops were estimated. Crop yields are assumed to follow the beta distribution and the parameters of the distribution are estimated using Maximum Likelihood Estimations. There are 300 (25 X 3 X 4) yield observations available for each crop. A total of ten production functions are estimated for each crop under each of the four rotation, seven for corn and three for soybean. We do not take into account the case of a continuous soybean rotation because it is not profitable and is never observed in practice.

Several studies have rejected the normality assumption on crop yields. Some of these authors, for example, [19], [12] and [71], agree that the crop yields are generally skewed, however they fail to reach a consensus on the nature of skewness. On the other hand [54] fail to reject normality of yield distribution for Kansas farm-level wheat, corn and sorghum. Using a Kolmogorov-Smirnov test for normality we can reject the hypotheses that corn and soybean yields follow the standard normal distribution at 5% significance level. It is assumed that a beta distribution is appropriate for yield function. Following [71] and [4], the parameters of the beta distribution function are conditioned on variables influencing yield. Ideally one would estimate a two-parameter beta distribution function with the following probability density function:

$$\beta(y|p,q,a,b) = \frac{\Gamma[p+q](y-a)^{p-1}(b-y)^{q-1}}{\Gamma[p]\Gamma[q]b^{p+q-1}} for\ a \leq y \leq b. \tag{2.9}$$

Parameters a and b in the above equation are bounds of support on the β distribution, Γ denotes the gamma function, and p and q are parameters which influence the shape of the probability density function. If we assume a production technology where the yield plateau shifts over time, then we have a time conditioned density function, given by $a = 0$ and $b = b_0 + b_1 t$. The parameters p and q are ideally functions of nitrogen application N, time t and the interaction terms. Under assumptions of concavity which requires specification of a second-order fit for the parameters of the β distribution function, one would require an estimation of 10 unknown parameters. There is not enough variation in the data set to estimate so many parameters. The problem is then simplified by excluding the nitrogen and time interaction terms and specifying the maximum attainable yield, b, exogenously. Given data restrictions, the following functional forms for p and q with 8 unknown parameters are estimated:

$$\begin{aligned} p &= p_1 + p_2 t + p_3 N + p_4 N^2 \\ q &= q_1 + q_2 t + q_3 N + q_4 N^2 \end{aligned}$$

Parametric estimates of production technologies are obtained with corn yields as functions of nitrogen application and time and soybean yields as function of time. The average yields from the data and the coefficient estimates of the above equations are presented in Table 2.7. The average increase in corn yield following soybean in CS rotation is 28.7% higher over continuous corn.

These estimated distribution functions are then adjusted to take into account the productivity differences across the counties. A county-wise shift parameter is constructed for corn and soybean by taking the ratio of the five year average yield of each county and the five year average yield of Floyd county. The construction of the shift parameter is based on the assumption that the experimental field is representative of the farms in Floyd county in terms of land quality and productivity. Communication with the data providers supported this assumption.

Thus the estimated distribution functions are assumed to hold true for Floyd county. A separate beta pdf for each of the remaining 98 counties in Iowa is obtained by parallel shifts of the beta pdf for Floyd county by the amount of the shift parameter.

In a given year the product of the price of a crop and its yield that year would give an internally consistent revenue for that year. The data for the corn and soybean prices were obtained from Chicago Board of Trade (CBOT) reports of futures prices. [32] and [53] argued in favor of using futures prices acreage response analysis on grounds of rational expectation and forecasting accuracy. The average settlement price observed during the first few weeks in April for the December and November maturity contract of the same year is considered to affect the planting decision of corn and soybean acreage, respectively. The annual crop prices are used to estimate the correlation structure between the yields and the crop prices.

For simulation purposes, it is assumed that crop prices follow the log-normal distribution with parameters explained in later sections. The assumption of log-normal distribution removes the possibility of negative prices and is common in agricultural economics literature ([40], [36]). The draws from log-normal distribution were adjusted to take into account price difference across counties. Data on the historical average annual basis for each county were obtained from the Center for Agricultural and Rural Development, Iowa State University. The basis describes the variation between the spot price of a commodity and the relative price of its futures contract. Differences in basis across counties arise due to transportation or other transaction cost differential causing a spatial price variation. The futures price for each commodity is adjusted using a basis differential. The historical average basis at maturity, reached during the month of October for each county is subtracted from that of the Floyd county to arrive at the basis differential.

Copula approach for correlated random variables

Literature on crop insurance programs not only emphasize the correlation between prices and yields but also correlations between yields from different crops. The Midwest corn producing states show strong negative corn and soybean yield-price correlation, forming a natural hedge with moderate revenue variability. This is characteristic of the Corn-belt region, where most farm-level yields are closely related to area-wide production and the areas production accounts for a significant share of world production. Accounting for yield correlation across crops in order to calculate more accurate insurance rates is also gaining popularity. [34] demonstrated that yield performance tends to be highly correlated across some crops and [35] show that corn and soybean specifically, are highly correlated.

Both types of correlation discussed above are taken into account during the simulation exercise. The estimated marginal distributions of the yield and price function and the correlation matrix obtained from the time series data are then used to estimate the joint distribution between the corresponding variables for each rotation. Again with our example of CCCS rotation, we would require an estimation of joint distribution of four crops and two prices. Under the Monte Carlo simulation approach, we need to draw from this joint distribution in order to calculate the profits for each rotation, given parameter values of nitrogen levels, price of nitrogen and time. Two procedures are generally used in agricultural economics literature to obtain random samples from correlated random variables. One commonly used method in constructing the rates for crop revenue insurance is the [50] method where uncorrelated random draws are combined using a weighted linear combination method. A limitation of the approach is the strict parametric specification leading to high sensitivity of these rates to correlations between prices and yields, which might not be measured accurately. Also there is no evidence that higher order cross moments are matched in this method. This method works well when the number of marginal distributions is small. A second method involves re-sorting as outlined by [46]. This method is used in many studies on crop insurance [39], [76]). In this method, correlation between independent random numbers could be obtained by orthogonal transformations. [47] note that the method is a close approximate, i.e., the rank correlation matrix of the distribution from which the draws are made is close to the prescribed rank correlation. Also, a multivariate distribution is not uniquely defined by its marginal distributions and its rank correlation.

In order to account for dependence between random variables we use an alternative procedure, the copula method. This a well-known tool in financial literature and its potential is recently being recognized in agricultural and environment fields, for example in [81]. This method is an improvement over the Johnson and Tenenbein approach since it can be easily implemented for more number of marginal distributions. [38] shows the similarity of results between IC method and the normal copula approach. The main advantage of the copula method is the flexibility it provides in addressing the correlation structure. This is done by specifying various functional form of the copula. [16] provide a comprehensive discussion of the various copula functions.

Definition and properties of copula function

A copula is a function that joins a multivariate probability distribution to a collection of univariate marginal probability functions. It is essentially a multivariate cumulative distribution function (cdf) with uniform pdfs. Consider n random variables $X_1, ...X_n$, with the following marginal distributions: $u_1 = F_1(X_1 \leq x_1), ...u_n = F_n(X_n \leq x_n)$. The multivariate cdf is the probability $P(X_1 \leq x_1, ..., X_n \leq x_n)$ or $F(X_1, ...X_n)$. Application of the method of copula is established upon the Sklar's Existence Theorem, which states that, "Given a joint distribution function and the respective marginal distribution functions, there exists a copula that binds the marginals to the joint distribution." [69]

$$F(X_1, ...X_n) = C(u_1, ...u_n) \tag{2.10}$$

where $C(.)$ is the joint distribution function of $U_j, j = 1...n$, which are correlated uniform random variables.

The copula provides information on the nature of dependence between different random variables, while the random variables themselves can follow any pre-specified distribution. In this approach the information on the individual marginal distribution functions and the dependence information are effectively separated from each other. Linear correlation is suitable as a measure of dependence when the underlying random variable is normal. However, when nonlinear transformations are applied to those random variables, linear correlation is no longer appropriate. Correlation measures for other distribution becomes more complicating. For example, for elliptical distributions a linear correlation estimator such as the Pearson product-moment correlation estimator has a very bad performance for heavier tailed data [63]. By separating the marginal distributions and the correlation structure, the copula approach helps in overcoming some of the problems of correlation as a measure of dependency. There is a very large number of copulas that can be specified. In this paper, Gaussian copulas are very effectively used to obtain the joint distributions.

Gaussian Copula

A Gaussian copula is constructed from the multivariate normal distribution as shown in the following equation:

$$
\begin{aligned}
C(u_1, ...u_n) &= \Phi_{u_1,...u_n,\Omega}(\Phi^{-1}(u_1), ..., \Phi^{-1}(u_n)) && (2.11)\\
&= \frac{1}{|\Omega|^{\frac{1}{2}}} exp\Big(- \frac{1}{2}\xi^T(\Omega^{-1} - I)\xi\Big) && (2.12)
\end{aligned}
$$

where Φ is the standard univariate normal cdf, $\Phi_{u_1,...u_n,\Omega}$ is multivariate normal cdf, Ω is an $n \times n$ correlation matrix and $\xi = \Phi^{-1}(u_j)$. Thus the copula can be separated into a marginal model for the inverse normal score $\Phi^{-1}(G_j(x_j))$ and joint distribution model of the inverse normal scores $\Phi_{u_1,...u_n,\Omega}(\Phi^{-1}(G_1(x_1)), ..., \Phi^{-1}(G_n(x_n)))$, where $G_j(x_j) = u_j$ is the cdf of x_j. The dependence structure is normal but the marginals can follow any distribution denoted by G.

The construction of Gaussian copulas require specification of the correlation structure among the variables. The Pearson correlation coefficient could be affected by the change of scale in the marginal variables. Instead, a rank correlation coefficient, such as Kendall's tau or Spearman's rho, is more appropriate. In this paper, Kendall's τ, is calculated between yields and prices for each rotation. It is invariant under strictly increasing transformation of the random variable.

The advantage of using the normal dependent structure lies in its analytical simplicity. It requires the estimation of only the correlation matrix. However, Gaussian copulas are not suitable when there is tail dependence among the correlated variables ([26]). Tail dependence refers to the probability that the outliers, positive or negative, occur jointly. For example, in finance, the Gaussian copulas are not suitable in modeling default risks which are considered to be rare events.

Correlated random draws

The procedure for drawing correlated random numbers is given by the following:

1. In the first step, k correlated normal random variables are generated 10,000 times for each rotation in each county. Here, k is the number of correlated marginals for each rotation. Correlation between the k variables are estimated from the data using Kendall's τ. For example, in case of continuous corn, random deviates are from the bivariate normal distribution since there are two marginal distributions, one for yield and the other for price. For CCCS, k is 6, three for the three years of corn yield, one for soybean yield and two for the prices of the two crops.

2. The normal cdf of each of these normal random variables are calculated to obtain random variables that are uniform over the interval [0,1] . These uniform variables define the dependence structure. Figure 2.6 shows 1,000 correlated random uniform variables for a CCS rotation. Kendall's τ is stated for each pair of correlated uniform random variates. First year corn and soybean yields have a higher correlation than the second year corn and soybean yields.

3. The marginal distributions are then constructed by taking the inverse of these uniform random variables. Crop yields are obtained by taking the inverse of the beta distribution with parameters given in Table 2.7 and specified levels of t and N. We use the built in functions in MATLAB to obtain the inverse of the beta distribution. For prices, the inverse of the log normal distribution is taken with mean and standard deviation equal to that of the futures prices.

Estimation of the expected profits and acreage allocations.

Expected profits from each rotation can now be calculated for each county. Monte Carlo simulations and grid search are useful techniques for extending the farm level analysis to the county. Nitrogen levels are varied from 0 to 300 lbs, by 1 lb increments. For a particular year, $\hat{t}$ and a particular nitrogen price $\hat{w}$ per lb, the expected profits per acre for each rotation is given by :

$$\begin{aligned}
\Pi_C &= Corn(t,N)P_{corn} - \hat{w}N \\
\Pi_{CS} &= 0.5Corn(t,N)P_{corn} + 0.5Soybean(t)P_{soybean} - 0.5\hat{w}N \\
\Pi_{CCS} &= 0.33Corn1(t,N1)P_{corn} + 0.33Corn2(t,N2)P_{corn} + 0.33Soybean(t)P_{soybean} \\
&- 0.33\hat{w}(N1+N2) \\
\Pi_{CCCS} &= 0.25Corn1(t,N1)P_{corn} + 0.25Corn2(t,N2)P_{corn} \\
&+ 0.25Corn3(t,N3)P_{corn} + 0.25Soybean(t)P_{soybean} - 0.25\hat{w}(N1+N2+N3)
\end{aligned}$$

where Corn1, Corn2 and Corn3 are the first, second and third year corn yields in a rotation and N1, N2 and N3 are the respective nitrogen application. P_{corn} and $P_{soybean}$ are the corn and soybean prices per unit.

We then perform a grid search over the different levels of nitrogen to arrive at the maximum profits and the profit maximizing nitrogen level. The maximum profits are compared across four rotation for each county to obtain the two most profitable rotations for each county. The ratio of the number of times the two highest profits are obtained gives the odds of occurrence of these profits. We apply this ratio as a smoothing factor to obtain the acreage under the two crops in all the 99 counties in Iowa. Under the assumption that all crop producers adopt the optimal rotation given the output and the input prices, the acres of cropland devoted to corn and soybean can then be estimated according to the rotation. For example, if the two most profitable rotations are CS and CCCS and they occur in the ratio p in the simulations then the total land under corn is $p * 50\% + (1-p) * 75\%$ and the total land under soybean is $p * 50\% + (1-p) * 25\%$. This is because if the optimal rotation is CS, then exactly 50% of cropland in a county is under corn production and 50% under soybean production. Similarly if the optimal rotation is CCCS for a given market condition then the entire cropland in the county is divided into 75% corn production and 25% soybean production. The percentage of corn and soybean acres in total corn and soybean acres obtained from the simulations are then transformed into percentage of corn and soybean acres in total acres in agricultural for each county.

The optimal distribution of cropland acreage between corn and soybean, as obtained from the optimal rotation is now a function of the nitrogen levels and the price of nitrogen. We can then study the effect of policies such as a corn-soybean rotation subsidy, a fertilizer tax and a nitrogen quota on ecological variables by varying the acreage allocation in each county.

Policy simulations

In order to do policy analysis, we first obtain the simulation results for profit maximizing nitrogen levels and maximum profits for each rotation and provide a prediction for the percentage of cropping acreage under corn and soybean for each county. A baseline is first established to evaluate the impacts of the policies. The nitrogen price did not change much towards the end of the field study period, that is, from 2001 to 2003. For the simulation, the nitrogen price is fixed at $0.21 per lb, the price for the given years ([22]). For corn prices, the log-normal distribution has a mean of $2.30 and a standard deviation of 0.20; for soybean prices they are $6.00 and 0.15, respectively.

The environmental consequences of ethanol industry is widely studied ([84], [85]). [2] estimate the amount of subsidy required to switch from row crop production to energy crop like switchgrass. Land under switchgrass production resulted in large reductions in nutrients in water. However, very little is known about what factors affect switchgrass production and a very large subsidy is needed for the conversion. [86] examine the impacts of higher crop prices on Iowa land going out of Conservation Reserve Program.

In this paper the ethanol effect is captured through rising corn prices, leading to a switch toward more corn intensive rotation like the CCCS rotation. Another high corn price baseline is set to study the effects of higher corn price due to increased demand for corn from the ethanol industry. Corn prices are log-normal with a mean of $3.75 and volatility of 27%. For soybean, the mean price is $7.00 and price volatility is 20%. These price levels and volatilities are based on the settlement of the December 2007 corn futures and November 2007 soybean futures contracts on December 21, 2006, [76]. The nitrogen price is assumed to be $0.31 per lb., from ISU Extension Publications.

The effect of ethanol can be seen in Table 2.8. The table provides descriptive statistics for the profit maximizing nitrogen levels, maximum profits and acreage allocation across the 99 counties in Iowa. In the low corn price scenario, CS rotation gives the highest profit followed by CCS rotation. In the high corn price scenario the highest profit is obtained from CCS rotation followed by CCCS rotation. This result is similar to [5], where at $6.00/bu price of soybean and $0.30 per pound of nitrogen, the break even corn price needed to induce a move away from CS rotation is $3.39 per bushel. This study shows a dramatic increase in corn acres from 56.56% to almost 70% of cropland, resulting in large changes in the ecological model. It is important to note that the percentage acres here are computed as a percentage of land under corn and soybean production alone.

Profit maximization and acreage allocation

The impact of different policies on acreage allocation is evaluated by taking the baseline values for the year 2001 for the two price scenarios. A subsidy given to the producers who adopt the corn-soybean rotation can be simulated by increasing the per acre revenues by a lump-sum amount. Similarly, the effect of a fertilizer tax can be simulated by increasing the fertilizer price. And the effect of a quota on nitrogen use can be simulated by specifying a smaller grid size of nitrogen input over which the profit functions are optimized. The effect of these policies on the average of the maximum expected profit, the average profit maximizing nitrogen levels, and the average expected corn and soybean acreage across all counties are presented in the Tables 2.9, 2.10 and 2.11.

Table 2.9 presents the effect of the three levels of fertilizer taxes at 15%, 45% and 90%

. A high fertilizer tax of 90% at the low price baseline scenario has a very small effect on the percentage of corn acres under production. One explanation for this could be that the Nitrogen fertilizer price in Iowa is cheap relative to its effect on output, leading to inelastic own-price effect of nitrogen fertilizer. [90] obtained the similar result that the elasticity of corn and soybean with respect to a fertilizer tax is very small. Under high prices, an input tax policy first corrects the increase in corn acreage due to the ethanol affect. A high tax (more than 90%) switches the acreage allocation under corn from as high as 70% to the low price scenario baseline. Beyond this level under high price scenario, the input tax policy acts similar to the low price scenario baseline and does not have a large effect on the acreage allocation.

The effect of a corn-soybean rotation subsidy, shown in Table 2.10 is the largest, especially under the low corn price baseline scenario. A $90.00 per acre subsidy given to producers who adopt the CS rotation under low price scenario decrease corn acreage by almost 7%. A low subsidy of only $18 per acre, comparable to more than a 100% fertilizer tax, corrects the distortion in the acreage allocation created by the ethanol effect. At this subsidy level the expected profit from CS rotation in each county is higher than that from CCS rotation, resulting in a decrease in corn acreage effectively because of the switch to the CS rotation. It should be noted that the percentage of corn acreage presented in the tables are a percentage of corn and soybean acres only. The percentage change in corn acreage as a percentage of total agricultural land would be smaller than these reported percentages. [103] estimate the corn-soybean rotation subsidy and find that at least $25.00 per acre is required before any change in rotation occurs and at this payment level there is an increase of only 1% of CS rotation in total acreage. Their conservative estimates could be attributed to the fact that their study is based on the entire UMRB region which includes other crop production systems other than corn and soybean. The way the rotation decisions are modeled here differs from their paper since they do not account for a more corn intensive rotation like CCS or CCCS rotation.

Tables 2.11 presents the effects of a restriction on nitrogen use per acre on acreage allocation and profits. The per acre change in the profit level gives an idea of the amount the farmers needs to be paid in order to compensate him for the profit loss due to a quota on nitrogen application levels. This, in effect, provides an estimation of the 'green payments' that need to be made to the farmers under restricted fertilizer use. For a nitrogen quota restriction from 200 lb per acre to 140 lb per acre and low corn prices, the average decrease in the profits is $3.93 per acre for the corn-soybean rotation, $9.42 per acre for the corn-corn-soybean rotation and $15.28 for the corn-corn-corn-soybean rotation. The corresponding profit changes under high corn prices are $7.16, $16.48 and $26 per acre respectively. The cost of any green payments associated with the policy increases by more than 100% under high corn prices. Under a high corn price scenario, a quota of 140 lbs per acre reduces the corn acreage by almost 11%. At this level of nitrogen, the corn-soybean rotation is the dominant rotation since the yield from corn-soybean rotation is the highest.

Figures 2.7 show the effect of the three policies on the acreage allocation under a high corn price scenario. The change in acreage allocation shown by the difference between the Figures 5E and 5A captures the ethanol effect. Except for few counties in the south-western region in Iowa, all the counties show that 10% of the county acres are converted from soybean to corn production due to the increase in the corn prices. The effectiveness of the three different policies in reversing the ethanol effect, i.e. moving from the high corn price baseline scenario in Figure 5A to the low price baseline scenario in Figure 5E, can be compared from Figures 5B, 5C and 5D. The change in the corn acreage allocation has important implications for ecological impacts, which will be studied in the next section.

Ecological effects

The acreage allocation obtained under different policies are then fed into the ecological models to obtain the percentage change in the environmental variables corresponding to a particular policy. The baseline values for ecological simulations are taken to be those of year 2001. The results for the changes in lake variables, river variables and pheasant population for changes in fertilizer taxes are shown in Tables 2.12, 2.13 and 2.14. For a fertilizer tax of 45%, the mean nitrogen levels in lakes decrease by -0.02% and the mean Secchi Depth increases by 0.80% under low corn price scenario. Given that some lakes in Iowa have as much as 16 mg/liter of N and fail to meet the maximum load criteria set by EPA, USDA initiative under the Total Maximum Daily Load (TMDL) program, a fertilizer tax would do little to rectify the situation. The input price tax performs better under high corn price baseline scenario. A tax of 45% decrease nitrogen levels by 0.31%. The effect on lake clarity is more dramatic than the effect on lake nutrients. The total suspended solids decrease by 3.5% and the Secchi Depth increases by 15% for the same level of tax, increasing the recreational value of the lakes. The effect of fertilizer taxes on river water nutrients is stronger. For river water quality, a 45% tax decreases nitrogen by more than 6% and phosphorus by almost 12%. Under a sufficiently high tax rate, almost 90%, the river water quality of that of the low price scenario is obtained. In the case of pheasant population, at 45% tax rate, there is a decrease by 0.24% under the low corn prices, and an increase by 2.49% under the high corn prices. The reverse effect of tax policy can be explained by the quadratic nature of the supply curve of pheasants. Under low corn prices, the tax policy improves water quality but lowers pheasant population. On the other hand in the situation of a high corn price, the tax policy is very effective for both water quality and pheasant population.

The effects of a nitrogen quota on the ecological variables are similar to those of a fertilizer tax. These are shown in Tables 2.15, 2.16 and 2.17. The nutrient levels in lakes change vary little under both high and low price scenarios. There is a marked improvement in lake clarity, as shown by more than 50% increase in Secchi Depth.

In terms of water quality, the corn-soybean rotation subsidy performs well as seen in Tables 2.18 and 2.19. There is a decrease in nitrogen levels in lakes by 0.19% and increase in Secchi Depth by more than 6% with a subsidy of $25 per acre, under low price scenario. Under high corn prices, a subsidy of $25 per acre, achieves the low corn price baseline water quality levels. River nitrogen decreases by more than 10% and phosphorus decreases by almost 15%. The subsidy is more effective in reducing river water pollution under high corn prices. Even though water quality improves with the subsidy, the pheasant population first increases and then decreases with the subsidy. A subsidy of $25 per acre increases pheasant population by 0.65% and obtains the low corn price baseline population.

Conclusions

This paper develops an integrated framework to quantitatively analyze the gains in environmental outcomes of alternative environmental policies that seek to reduce non-point pollution. It differs from earlier studies in attempting to address multiple environmental outputs. The framework developed in this paper explicitly incorporates the economics of crop rotation decisions leading to changes in acreage allocation between crops.

Several important insights are obtained from this study. First, each policy is different in its implication on different environmental outcomes. For example, a policy like a CS rotation subsidy to induce more land under nitrogen fixing soybean can have substantial effect on water quality but detrimental effects on pheasant population. Second, the commodity market price scenario should be taken into account while formulating an environmental policy. Different environmental policies have different effects under different commodity price scenario. The CS rotation subsidy is beneficial to pheasant population under high price scenario, but detrimental under low prices. Finally, under high corn prices the environmental damages are increased and the environmental policies become more expensive. However, there are threshold levels of policies, under high price scenarios, up to which point the policies are highly effective. Beyond these levels, once the acreage allocation under low corn prices are achieved, the policies have marginal effects. Since acreage allocation is modeled on the profitability of rotations, the policies can correct the acreage allocation resulting from a market of high corn prices. With suitable levels of taxes, subsidies or quotas, the corn-soybean rotation can be achieved as the dominant rotation with most profits, leading to low price baseline levels of environment quality. However, these policy instruments are not very effective in improving environmental quality beyond these low price baseline scenarios.

The results of this study provide information for future policy debates at the WTO forum. From the WTO perspective, a fertilizer tax and nitrogen quota based on a fixed, historical base period is acceptable since it is believed that they do not change the acreage allocation [9]. As these results show, fertilizer taxes and quotas can have substantial effect on acreage allocation under high commodity prices. On the other hand, corn-soybean rotation subsidy is more likely to be challenged at the forum or could come in conflict with domestic bio-fuel promotion policies, some of which might have the converse effect on corn acreage allocations. In such scenarios, the subsidy could be justified due to its dramatic effect on reducing water pollution.

The analysis of the ecological models in this paper is limited by data availability. The model could be extended to include other ecological effects of agricultural production, such as carbon sequestration. A desirable extension of this study would involve formulating an all-encompassing environmental index giving proper weights to different benefits.

Table 2.1 Summary Statistics for Ecological variables

Variables	N	Mean	Std.Dev	Min	Max
Secchi Depth (m)	926	1.18	0.88	0.05	8.1
Lake Nitrogen (mg/l)	926	2.54	2.85	0.17	16.75
Lake Phosphorus (μg/l)	926	121.13	96.44	14	760
Lake TSS (mg/l)	831	19.65	21.82	2	218
River Nitrogen	453	5.45	2.89	0.07	14.23
River Phosphorus	453	105.7	105.5	0	770
Pheasant	1971	37.47	29.70	0	220

Table 2.2 Lake Water Quality Estimation

	Nitrogen		Phosphorus		TSS		Sechhi Depth	
Variable	Estimate	t value	Estimate	t value	Estimate	t value	Estimate	t value
Intercept	2.37	1.00	-306.10	-2.93	-33.54	-1.53	5.97	6.31
Corn Area	7.22E-06	1.21	1.22E-04	0.46	9.70E-05	1.73	-7.25E-06	-3.03
Soybean Area	6.04E-06	0.88	1.96E-04	0.65	-4.27E-06	-0.07	6.24E-06	2.28
Watershed Area	7.40E-05	10.75	1.30E-04	0.43	1.79E-04	2.80	-2.00E-05	-7.27
Volume of lake	-6.42E-09	-5.25	-2.27E-08	-0.42	-4.05E-08	-3.56	4.75E-09	9.69
Slope	-0.28	-2.99	-8.60	-2.05	-2.27	-2.58	0.15	3.87
CSR†	-0.05	-2.24	-2.01	-2.04	-0.56	-2.70	0.02	1.79
Kfactor	-2.23	-0.52	-49.69	-0.26	-56.52	-1.42	4.11	2.40
Precipitation	0.05	2.99	0.00	0.00	0.17	1.14	-0.01	-1.77
Temperature	-0.27	-5.10	6.03	2.61	-0.79	-1.63	-0.09	-4.27
SDR††	0.39	14.11	0.93	0.77	-0.84	-3.31	0.02	1.71
Clay	0.02	0.39	10.36	4.08	3.24	6.08	-0.14	-6.32
Sand	0.13	5.80	3.49	3.56	0.69	3.34	-0.01	-0.57
SoilP	2.18	5.28	55.82	3.06	14.37	3.75	-0.83	-5.03
Weighted R^2	0.2879							

† Corn Suitability Rating

†† Sediment Delivery Rate

Table 2.3 Lake Water Quality Estimation without Soybean Acres

	Nitrogen		Phosphorus		TSS		Sechhi Depth	
Variable	Estimate	t value	Estimate	t value	Estimate	t value	Estimate	t value
Intercept	2.73	1.17	-294.40	-2.86	-33.80	-1.57	6.34	6.79
Corn Area	1.20E-05	5.38	2.81E-04	2.84	9.30E-05	4.43	-2.18E-06	-2.44
Watershed Area	7.50E-05	10.92	1.42E-04	0.47	1.79E-04	2.80	-2.00E-05	-7.26
Volume	-6.47E-09	-5.29	-2.43E-08	-0.45	-4.05E-08	-3.57	4.69E-09	9.56
Slope	-0.29	-3.02	-8.70	-2.08	-2.27	-2.58	0.14	3.78
CSR†	-0.05	-2.28	-2.03	-2.07	-0.56	-2.70	0.02	1.70
Kfactor	-2.62	-0.62	-62.33	-0.33	-56.24	-1.42	3.70	2.17
Precipitation	0.05	2.92	-0.05	-0.07	0.17	1.15	-0.01	-1.99
Temperature	-0.27	-5.20	5.90	2.56	-0.79	-1.64	-0.09	-4.48
SDR††	0.39	14.17	0.97	0.80	-0.84	-3.32	0.02	1.82
Clay	0.03	0.51	10.59	4.21	3.23	6.13	-0.14	-6.05
Sand	0.12	5.78	3.32	3.52	0.69	3.49	-0.01	-1.21
SoilP	2.15	5.23	54.85	3.02	14.39	3.77	-0.86	-5.22
Weighted R^2	0.2857							

† Corn Suitability Rating

†† Sediment Delivery Rate

Table 2.4 River and Stream Water Quality Estimation.

	Nitrogen		Phosphorus	
	Estimate	t value	Estimate	t value
Intercept	5.60169	6.14	0.128798	3.67
Corn Area	3.10E-05	3.51	3.64E-07	1.06
Soybean Area	-3.00E-05	-2.82	-4.80E-07	-1.16
Flow	2.24E-04	1.99	3.91E-06	0.9
Slope	-0.38015	-0.81	-0.01899	-1.05
EI for water	0.37083	0.9	0.034312	2.18
EI for wind	-0.05768	-0.29	0.002964	0.38
Weighted R-square	0.0562			

Table 2.5 River and Stream Water Quality Estimation without Soybean Acres.

	Nitrogen		Phosphorus	
	Estimate	t value	Estimate	t value
Intercept	5.336681	5.83	0.12459	3.57
Corn Area	9.02E-06	2.18	7.91E-09	0.05
Flow	0.000209	1.85	3.68E-06	0.85
Slope	-0.40805	-0.86	-0.01943	-1.07
EI for water	-0.13608	-0.37	0.026262	1.85
EI for wind	-0.01916	-0.09	0.003576	0.46
Weighted R-square	0.0402			

Table 2.6 Estimation of pheasant supply.

Pheasant counts		
	Estimate	t value
Intercept	15.26	3.10
Corn Area	289.18	9.65
Soybean Area	-85.12	-12.13
Square of corn area	-373.08	-8.03
Population Density	-25.01	-6.16
Northwest Dummy	10.11	6.65
Precipitation	-0.18	-2.57
R square	0.0706	

Table 2.7 Mean values and parameter estimates of the crops under rotation.

Rotation	CS		CCS			CCCS				Continuous
	Corn	Soybean	Corn1	Corn2	Soybean	Corn1	Corn2	Corn3	Soybean	Corn
Mean	137.57	45.24	136.54	107.46	48.29	134.37	108.27	105.53	49.81	106.88
Maximum yield	212.95	68.60	216.60	208.44	72.62	216.90	201.70	201.00	75.90	212.60
p function										
Constant	7.04	4.17	8.72	5.21	6.62	10.14	6.17	6.69	6.93	5.28
N	0.03		0.02	0.04		0.01	0.04	0.04		0.07
t	-0.01	0.19	-0.12	-0.13	0.09	-0.12	-0.16	-0.14	0.13	-0.11
N^2	-1.62E-04		-8.01E-05	-1.26E-04		-9.14E-05	-1.67E-04	-1.54E-04		-2.49E-04
q function										
Constant	8.49	4.29	9.62	11.05	5.61	11.64	11.79	13.29	6.59	12.33
N	-0.04		-0.05	-0.05		-0.06	-0.06	-0.08		-0.04
t	-0.07	-0.08	-0.12	-0.12	-0.13	-0.14	-0.10	-0.08	-0.16	-0.11
N^2	8.16E-05		1.19E-04	1.11E-04		1.42E-04	1.30E-04	1.62E-04		4.82E-05

Table 2.8 The effect of ethanol on the mean acreage allocation.

	Continuous Corn		C in CS		CCS			CCCS				Percentage corn acres†
	N	Π‡	N	Π‡	N1‡	N2‡	Π‡	N1‡	N2‡	N3‡	Π‡	
						Low corn price						
Mean	208.10	278.73	186.19	333.27	180.02	197.79	325.24	181.69	201.25	201.47	317.32	57.56
Std.Dev.	1.95	26.87	2.2	24.6	1.4	1.5	25.8	1.5	1.1	1.0	26.1	0.28
Max	212	348.71	191	402.29	183	201	396.28	185	204	204	388.68	58.1
Min	203	219.31	181	277.98	176	194	269.3	178	199	199	260.46	56.79
						High corn price						
Mean	210.14	460.77	188.57	470.96	181.67	199.29	481.68	183.18	202.43	202.55	480.43	69.85
Std.Deviation	1.63	40.73	1.8	34.0	1.3	1.2	36.8	1.2	0.9	0.8	37.9	3.37
Max	213	561.25	192	561.98	184	202	578.04	186	204	204	578.42	70.9
Min	206	364.96	184	392.82	178	196	397.26	180	200	200	392.9	58.46

†Corn acres as a percentage of total corn and soybean acres

‡ Π are profits and N1, N2 and N3 are profit maximizing levels of N in the first, second and third years .

Table 2.9 The effect of fertilizer tax on the mean acreage allocation.

N tax	Continuous Corn		C in CS		CCS			CCCS				Percentage corn acres†
Percent	N	Π‡	N	Π‡	N1‡	N2‡	Π‡	N1‡	N2‡	N3‡	Π‡	
Low corn price												
15	204.60	272.13	182.33	330.32	177.51	195.30	321.30	179.08	199.42	199.91	312.74	57.47
45	197.16	259.47	174.00	324.71	317.41	172.17	190.20	173.79	195.48	196.31	303.73	57.26
90	188	241.51	160.34	316.84	163.76	182.23	302.42	165.21	189.12	190.56	290.61	56.96
High corn prices												
15	207.17	451.06	185.15	466.61	179.37	197.11	475.81	180.92	200.86	201.05	473.62	69.21
45	200.75	432.09	177.98	458.17	174.65	192.57	464.28	176.27	197.33	197.93	460.20	66.43
90	189.90	404.82	166.08	446.16	167.22	185.48	447.54	168.75	191.76	192.92	440.58	60.51

†Corn acres as a percentage of total corn and soybean acres

‡ Π are profits and N1, N2 and N3 are profit maximizing levels of N in the first, second and third years .

Table 2.10 Effect of corn-soybean subsidy on corn acres for two price scenarios.

Subsidy per acre	Low corn price % corn acres†	High corn price % corn acres†
0	57.56	69.85
10	56.63	64.58
25	55.30	57.42
50	53.44	55.90
75	51.98	54.50
90	51.07	53.72

†Corn acres as a percentage of total corn and soybean acres

Table 2.11 The effect of N quota on mean profits and mean percentage of corn acres.

N Quota	Continuous Corn π	CS π	CCS π	CCCS π	Percentage corn acres†
		Low corn price			
200	278.40	333.27‡	325.17‡	317.31††	57.56
170	272.77	332.72	323.38	313.24	57.44
140	261.18	329.34	315.75	302.03	57.02
		High corn prices			
200	459.93	470.96‡	481.68‡	480.34††	69.80
170	449.79	469.80	478.27	473.33	67.90
140	429.954	463.80	465.20	454.34	58.41

‡: quota is non-binding.

††: quota is non- binding for the first year corn only

†Corn acres as a percentage of total corn and soybean acres

Table 2.12 Effects of fertilizer tax on lake variables

	Lake N		Lake P		TSS		Secchi Depth	
N tax	Average	% change in	Average	% change	Average	% change	Average	% change
			State of low corn prices					
0	3.0566		114.04		20.75		1.0981	
15	3.0564	-0.00654	114.06	0.017538	20.732	-0.08675	1.1009	0.254986
45	3.0559	-0.0229	114.09	0.043844	20.687	-0.30361	1.1069	0.801384
90	3.0551	-0.04907	114.14	0.087689	20.618	-0.63614	1.1161	1.639195
			State of high corn prices					
0	3.0898		111.95		23.617		0.71666	
15	3.0869	-0.09386	112.13	0.160786	23.369	-1.05009	0.7496	4.596322
45	3.0802	-0.3107	112.56	0.544886	22.784	-3.52712	0.82751	15.46759
90	3.0649	-0.80588	113.52	1.402412	21.465	-9.11208	1.0033	39.99665

Table 2.13 Effects of fertilizer tax on river and stream variables.

N tax	Nitrogen Average	Nitrogen % change in	Phosphorus Average	Phosphorus % change in
State of low corn prices				
0	5.93		114.31	
15	5.706	-3.84	101.15	-11.51
45	5.54	-6.64	100.8	-11.82
90	5.51	-7.15	100.28	-12.27
State of high corn prices				
0	7.14		122.95	
15	7.14	0.00	122.89	-0.05
45	6.81	-4.62	118.32	-3.77
90	5.8	-18.77	104.52	-14.99

Table 2.14 Effects of fertilizer tax on pheasant population.

N tax	Average Pheasant	% change in pheasant population
State of low corn prices		
0	35.67	
15	35.64	-0.08
45	35.58	-0.24
90	35.49	-0.50
State of high corn prices		
0	35.38	
15	35.17	-0.59
45	36.26	2.49
90	35.86	1.36

Table 2.15 Effects of N quota on lake variables.

	Lake N		Lake P		TSS		Secchi Depth	
N tax	Average	% change	Average	% change	Average	% change		
State of low corn prices								
200	3.06	0	114.04	0	20.75	0.02	1.10	-0.009
170	3.06	-0.01	114.06	0.017	20.72	-0.12	1.10	0.34
140	3.05	-0.046	114.13	0.08	20.63	-0.57	1.11	1.47
State of high corn prices								
200	3.08	-0.27	111.90	0.04	23.62	0.001	0.72	0.001
170	3.08	-0.20	112.34	0.35	23.08	-2.28	0.79	10.01
140	3.06	-1.00	113.9	1.74	20.95	-11.30	1.07	49.58

Table 2.16 Effects of N quota on river variables.

	River N		River P	
N tax	Average	% change	Average	% change
State of low corn prices				
200	5.93	0.01	114.32	0.09
170	5.92	-0.29	114.08	-0.20
140	5.86	-1.27	113.27	-0.91
State of high corn prices				
200	7.12	0.03	123	0.04
170	7.47	4.64	135.57	10.26
140	6.05	-15.27	115.92	-5.72

Table 2.17 Effects of N quota on pheasant population.

	Pheasant population	
N tax	Average	% change
State of low corn prices		
200	35.67	0.01
170	35.63	-0.10
140	35.52	-0.41
State of high corn prices		
200	35.40	0.4
170	34.94	-1.24
140	35.88	1.42

Table 2.18 Effects of CS rotation subsidy on lake variables

Subsidy	Nitrogen		Phosphorus		TSS		Secchi Depth	
(per acre)	Average	% change	Average	% change	Average	% change	Average	% change
State of low corn prices								
0	3.0566		114.04		20.75		1.0981	
10	3.054	-0.08506	114.21	0.149071	20.526	-1.07952	1.1283	2.750205
25	3.0507	-0.19302	114.42	0.333216	20.242	-2.44819	1.1662	6.201621
50	3.0458	-0.35333	114.72	0.596282	19.822	-4.47229	1.2221	11.29223
75	3.042	-0.47765	114.96	0.806734	19.492	-6.06265	1.2661	15.29915
90	3.0396	-0.55617	115.12	0.947036	19.285	-7.06024	1.2936	17.80348
State of high corn prices								
0	3.0898		111.95		23.617		0.71666	
10	3.077	-0.41427	112.76	0.723537	22.512	-4.67883	0.86381	20.53275
25	3.0563	-1.08421	114.06	1.88477	20.727	-12.2369	1.1016	53.71306
50	3.0523	-1.21367	114.31	2.108084	20.382	-13.6978	1.1475	60.11777
90	3.0466	-1.39815	114.67	2.429656	19.89	-15.781	1.2131	69.27134

Table 2.19 Effects of CS rotation subsidy on river and stream variables

Subsidy (per acre)	River N Average	River N % change	River P Average	River P % change
State of low corn prices				
0	5.93		114.31	
10	5.46	-7.99	99.72	-12.76
25	5.3	-10.68	97.41	-14.78
50	5.07	-14.56	94.2	-17.59
75	4.88	-17.76	91.64	-19.83
90	4.76	-19.78	89.99	-21.28
State of high corn prices				
0	7.14		122.95	
10	6.66	-6.73	116.22	-5.47
25	5.56	-22.13	101.1	-17.77
50	5.38	-24.65	98.46	-19.92
90	5.1	-28.57	94.7	-22.98

Table 2.20 Effects of CS rotation subsidy on pheasant population

	Pheasant Population	
Subsidy (per acre)	Average	% change
State of low corn prices		
0	35.67	
10	35.68	0.04
25	34.94	-2.04
50	34.21	-4.09
90	33.11	-7.17
State of high corn prices		
0	35.38	
10	35.80	1.19
25	35.61	0.65
50	35.14	-0.68
90	34.32	-3.00

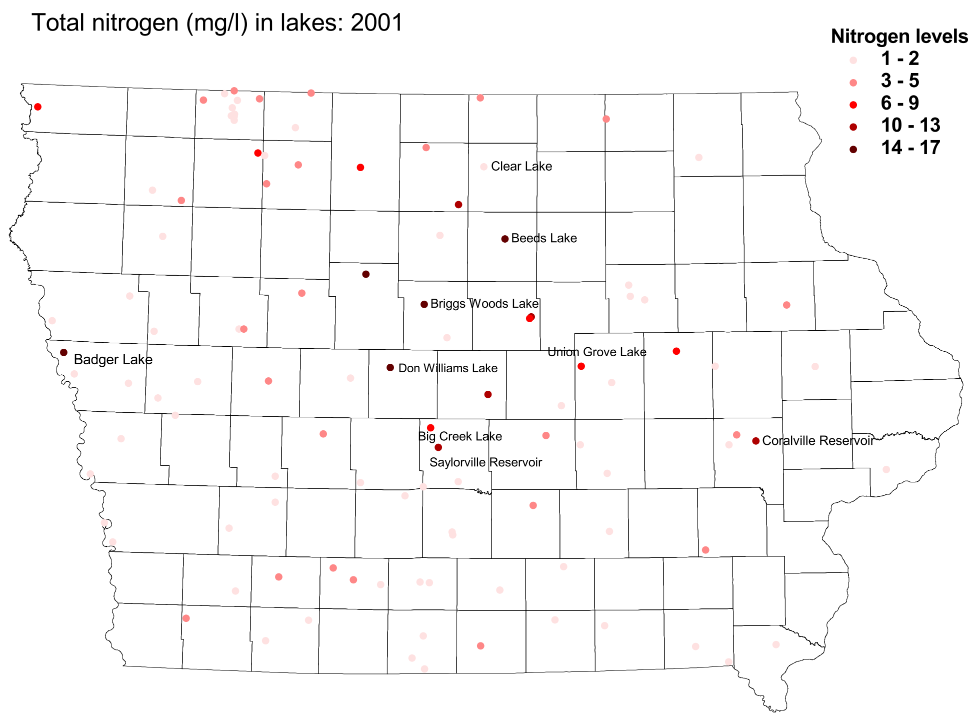


Figure 2.1 Nitrogen levels in lakes

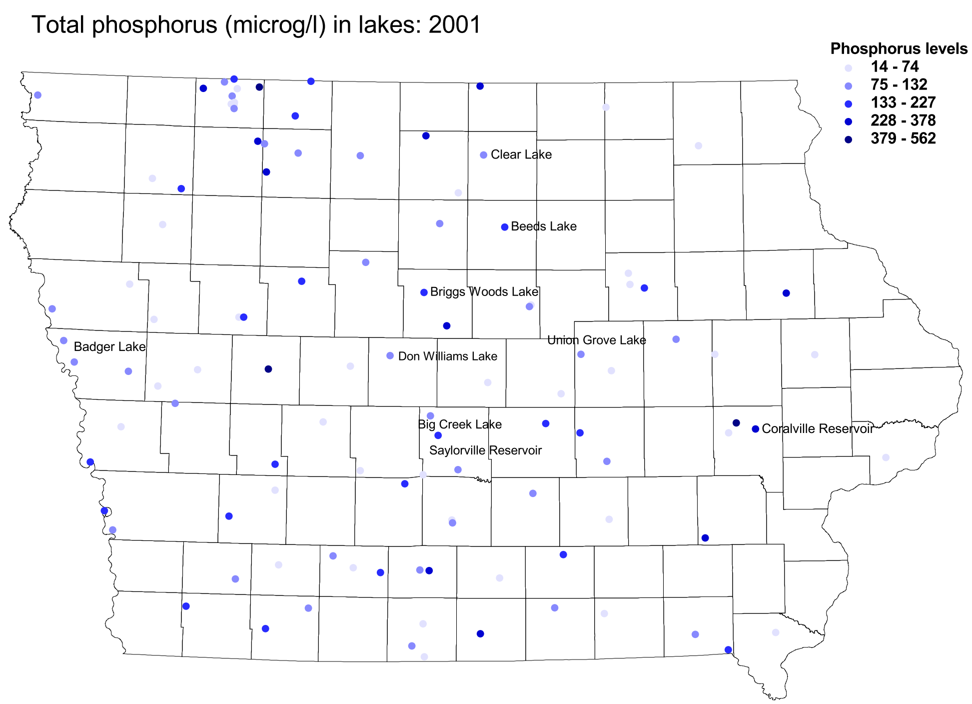


Figure 2.2 Phosphorus levels in lakes

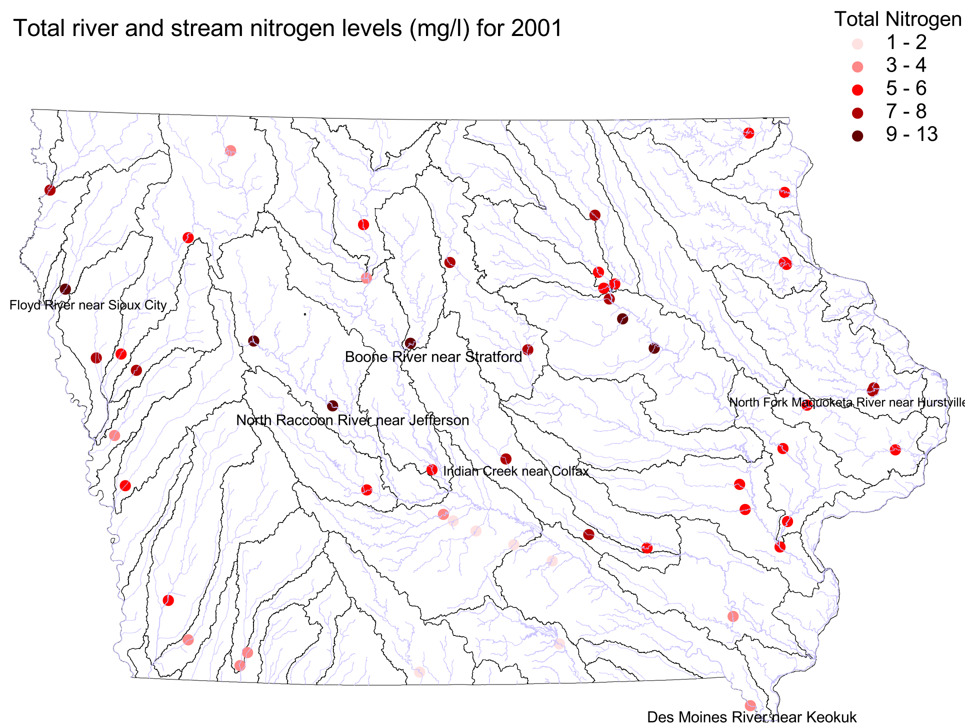


Figure 2.3 Nitrogen levels in rivers

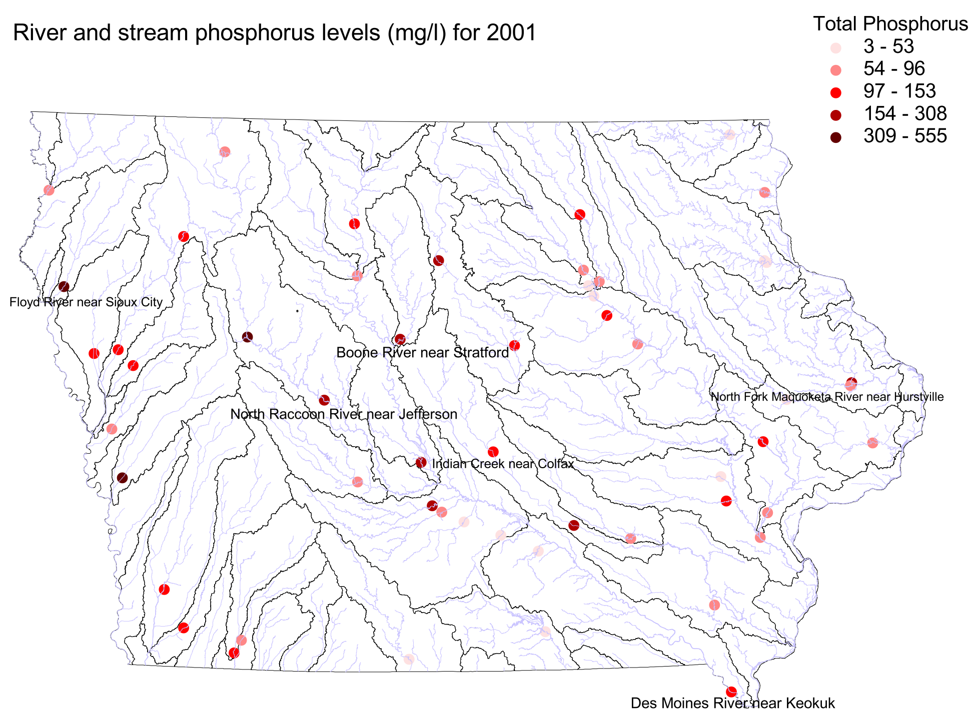
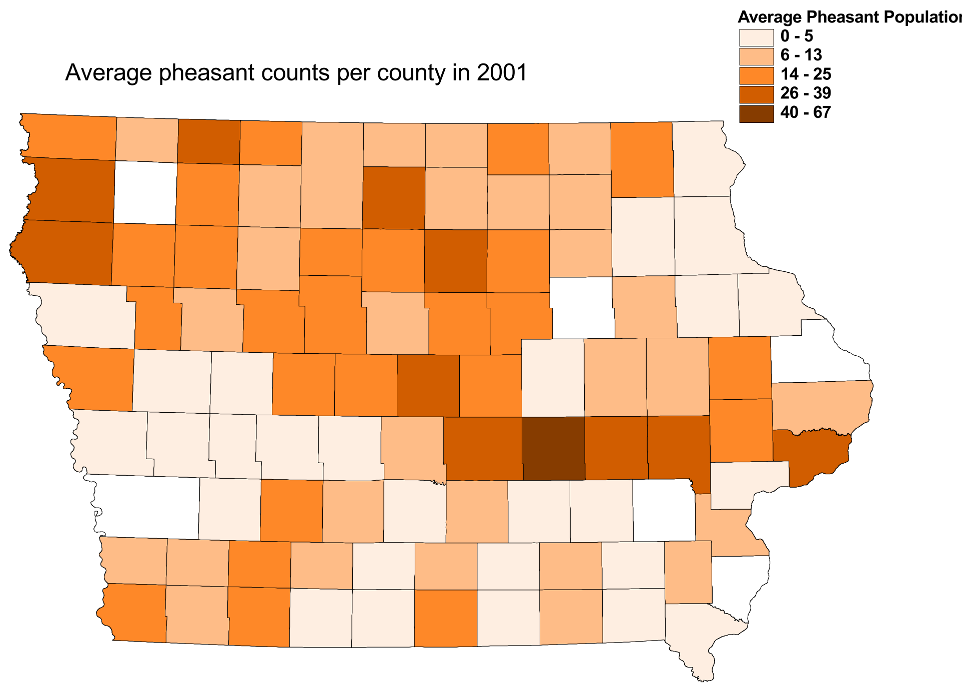


Figure 2.4 Phosphorus levels in rivers



No data available for the blank counties

Figure 2.5 Average population of ring necked pheasant per 30 mile route across Iowa counties in 2001

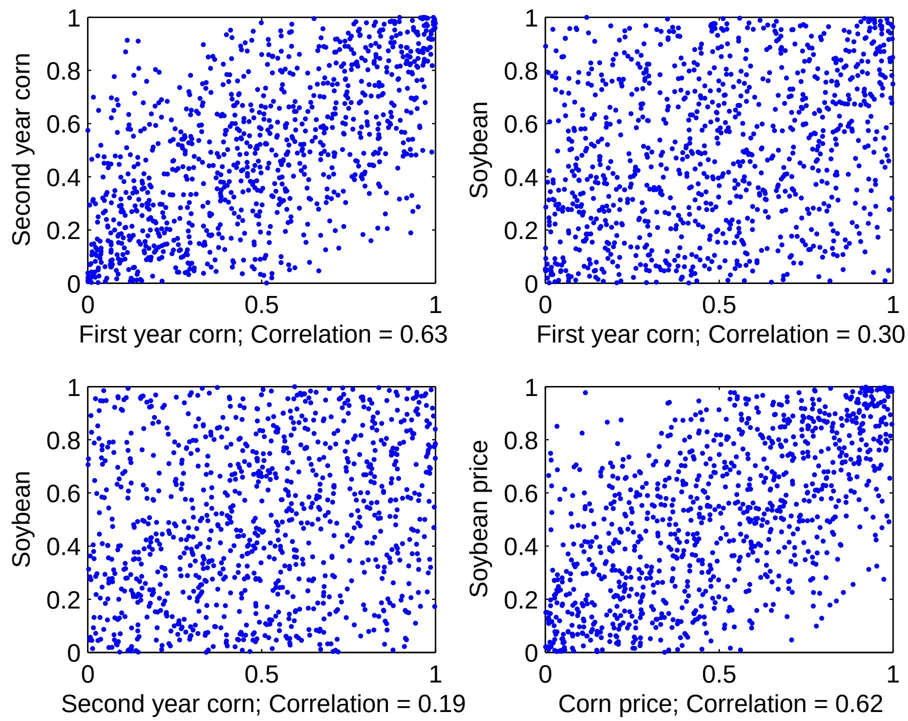


Figure 2.6 Simulated Dependent Variables for CCS Rotation

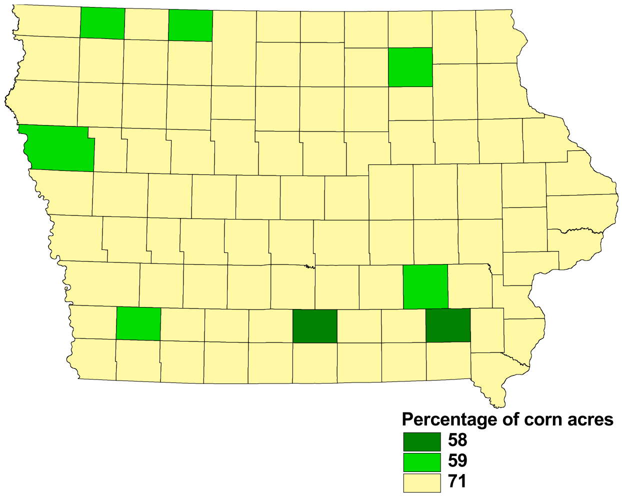


Figure 5A: High corn price baseline scenario

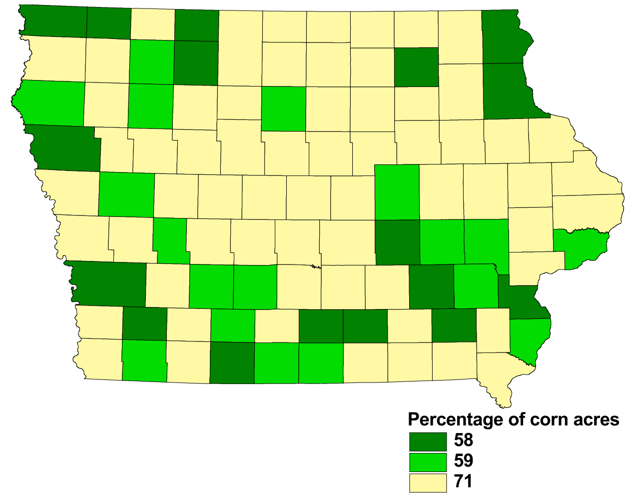


Figure 5B: Input tax of 45%

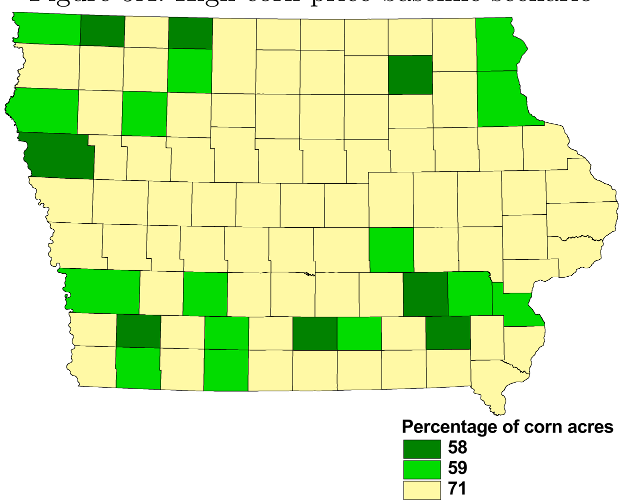


Figure 5C: Input use restriction, 170 lb of N per acre

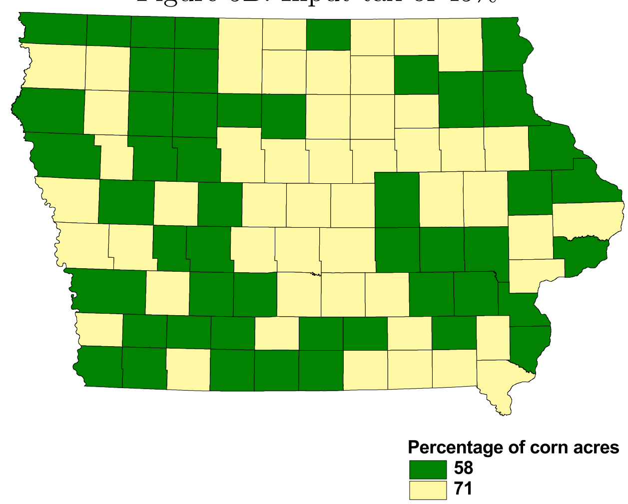


Figure 5D: CS rotation subsidy of $10 per acre

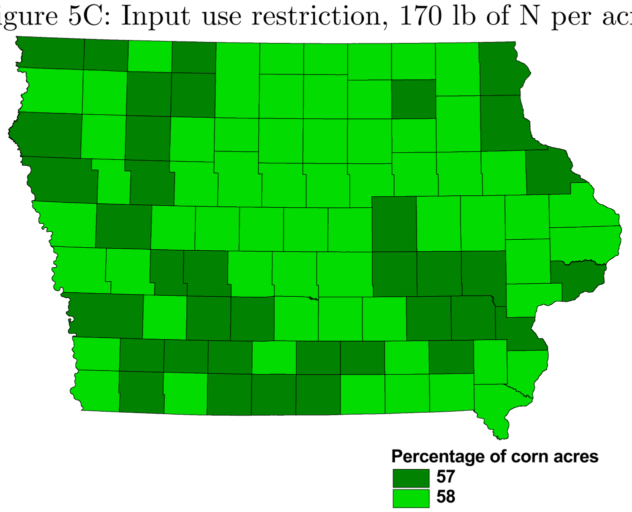


Figure 5E: Low corn price baseline scenario

Figure 2.7 Simulations under different policy scenarios

3. Further Investigation of the Effects of Meat Recalls and Correlations on Consumer Demand and Volatility.

Abstract

Product recalls provide information on new sources of contamination or defects that could trigger a series of recalls. Closely related US beef, pork and poultry data are examined for the presence of recall clusters. Specifically, the correlations between the recall events for these products are determined. There is reason to suspect the presence of pairwise correlation which can be attributed to a number of factors, including the use of similar technologies as meat packers operate across species. Also, the recall correlations vary over time. It is well established that the recalls convey important food safety information and are found to affect consumer demand. An important contribution of this paper is to examine whether consumers take into account the possibility that another recall is likely to occur soon. A static model is used to determine the change in the shares of expenditure for different recall correlation patterns. An absolute price version of the Rotterdam demand model is estimated. Recalls are assumed to follow a Poisson arrival process and copula techniques are used to generate correlated variables to simulate the effects of correlation on the shares of expenditure of the three food groups. It is found that although the mean values of the shares of expenditure remain unchanged in the simulation, the variance of the shares change with a change in the specified correlation structure. This suggests growing uncertainty with higher levels of correlation, and has implications for the demand elasticity estimates of recall. This study is further extended to test whether recalls increase the volatility of retail prices.

Introduction

Product recalls serve as essential policy tools to align social and private incentives with respect to product safety. The recent spurt in recalled products raises an important issue in underpinning the economic forces that lead to consecutive recall events. After the extensive Toyota recalls that hit the headlines since October 2009, auto makers worldwide have been quick to initiate recall process rather than wait for long, drawn out government inquiries. According to industry analysts, auto makers are now more aware of the harmful publicity that results from not addressing a safety problem quickly. More recently Iowa's Hillandale Farms recalled more than 170 million eggs after laboratory tests confirming salmonella. This occurred one week after another Iowa farm, Wright County Eggs, recalled 380 million eggs. Both the recalls are said to be related since both the plants could have a rodent problem, or both plants could have gotten hens that were already infected or feed that was contaminated. These examples motivate the need to establish correlation patterns between the recalls and the ramifications of these correlations on the economic agents. Understanding this recall correlation is essential in forecasting future recalls as it relates these events with some identifiable pattern. Public-private crisis management procedures could act on these relationships to improve recall effectiveness.

Usually when an outbreak is detected, there is heightened inspection, testing and investigation which brings to the fore safety measures that are violated or overlooked. Recall incidence are generally more prevalent during these times. The correlations could also arise because similar technologies are used during production, slaughter, processing, and distribution of hogs, poultry and beef. In the production phase, for example, hogs and poultry being non-ruminants and produced in confinement could be more closely related than cattle which are typically raised on range or pasture lands and then placed in a feedlot. According to [41], an important aspect of food quality failure is the interconnected stages and inputs in the food production systems. A [94] report found that larger feedlots had higher incidence of diseases, despite evidence that larger feedlots took more precautions. Larger feedlots supplied to larger packing companies. Technical progress and scale effects have led to oligopsony power in the meat packing industry

([31]). Further these few large companies, Tyson Foods Inc., Smithfield, Cargill and JBS now operate across species, with a myriad of value added/processed and packaged beef, pork and poultry products. Similar technology and company policy would affect safety issues for a large number of these products.

This paper borrows the idea of default correlation from financial literature, to analyze these events. Default correlation measures whether credit risky assets are more likely to default together or separately. The main hypothesis of this paper is that the recall events for related food groups might be correlated. A recall group event correlation is the likelihood that if there is a recall in a particular food group, there will be a recall in another food group soon after. That is, if a certain pork product is found to be contaminated then it is likely that a contaminated beef product could be detected in the near future. This paper then proceeds to show how the relationships between food recall events might affect the parameter estimates of two different models. Firstly, failure to account for the dynamic changes in the food recall correlation might produce inefficient parameter estimates of the consumer demand response to meat recall information. Secondly, the volatility of food prices might be affected in the presence of food recalls. Price volatility indicates the range within which prices might vary in future and this complicates the production, investment and consumption decisions of businesses and consumers.

Consumer demand response to meat recall information has been studied extensively. Some of the factors that act as meat demand shifters, for example, food safety and product recalls and related news have been included in demand estimations by [78], [67], [13]. In general it has been observed that when USDA Food Safety Inspection Service (FSIS) beef product recalls increase, beef demand declines. Moreover, beef product recalls have a significant positive spill-over effect on poultry demand, suggesting that consumers shift away from beef and toward poultry products in response to beef food safety recalls. Previous recall events are included to study the long run and short run effects of recall. However, there are no studies on the effects of the second order properties of recalls, that is correlation and autocorrelation. The recall events between several food groups like beef, pork and poultry may be correlated and this correlation might vary over time. In the presence of this correlation, the parameter and elasticity estimates from the consumer demand model may not be efficiently estimated. Consumers may face increased uncertainty over their consumption decisions if they anticipate more recalls in the future. Time varying correlations are extensively studied in stocks and bond markets ([11]), international stock prices ([64]) and play a significant role in pricing other financial instruments and portfolio management. In all these studies, the assumption of a constant correlation are shown to lead to biased or inefficient parameter estimates.

There are several studies on the effect of recall on stock prices ([98]), futures prices ([65]) and the stock market reaction to food recalls based on industry or firm structure ([82]). [98] study the effect of food recalls on the conditional variance of the stock returns and use the GARCH approach. They reject the assumption of a constant variance of the stock prices of two companies. Salin and Hooker (2001) investigated the stock market reaction to food recalls using an event study approach. They found statistically significant evidence of a negative effect on returns. [65] quantify the effects of beef and pork recall announcements on daily live cattle and lean hog futures market prices. Their results indicate that, in general, pork and beef recall events have not systematically impacted daily lean hog and live cattle futures market. A news impact curve is often used to measure how new price information is incorporated into volatility ([27]). Inclusion of the recall events in the volatility might help in explaining a part of the volatility and provide a more accurate measure of the conditional variance.

The remainder of the article proceeds as follows. First, we discuss the available data on recall and provide evidence for the presence of constant or dynamic correlations. Following that, we present two illustrations to study the effect of ignoring a time-varying structure. One develops the system of equation to study the demand for meat and provides simulated results on the parameter estimates in the presence of recall correlation. The other presents an asymmetric GARCH model, where negative shocks have a greater effect, to show how food recalls might affect the volatility of food prices. In the conclusion we revisit the question of including correlation structure and discuss substantive insights from doing so.

Food recalls

Recalls of meat products in the USA are regulated by the Federal Meat Inspection Act of 1906 and the Poultry Products Inspection Act of 1957. USDA-FSIS has responsibility for ensuring that meat and poultry are safe, and accurately labeled. But for the most part, recalls are voluntary action by firms to remove the contaminated product from market. Following, [67], the recall event measures are constructed by aggregating the total number of recalls per quarter for beef, pork and poultry. The recall dataset is obtained from the FSIS and comprises of data from 1994 to 2008. Figure 3.1 shows the number of recall events for beef, pork and poultry products cumulated per quarter. During the period 2000 to 2003, average quarterly recalls were high for all three food groups. Summary statistics of the number of FSIS recall events per quarter are provided in Table 3.1. The average recall per quarter for beef is 5.23, for pork it is 1.72 and for poultry it is 3.1. Beef recalls are most frequent, followed by poultry and then pork.

Correlation Analysis

A recall action is initiated based on the results of any one of the following actions: regular sampling tests by FSIS, microbial testing or product inspection by the firms, consumer complaint reports or other actions like results from epidemiological tests by CDC, reports from state health departments, FDA etc. Each of these agents has an interest in the timely discovery of the problem or contamination and effective handling of the recall process. The recall events of different industries could be correlated, if any of these agents internalize the information on a recent recall to detect problems. For example, recalls by other firms could lead companies to investigate their own production and supply chain processes to avoid similar hazards. That is, if the recall events are temporally correlated then it could be an outcome of search and information acquisition. As soon as a recall event occurs it provides immediate information on the possible occurrence of the next recall event. That is, the first few days following a recall event might increase the likelihood of recalls in the same food supply chain. As the number of days increase after a recall the possibility of another recall decreases as more information is obtained from the last recall in terms of learning about the pathogen, maintaining safe food handling techniques etc. The rest of this section analyzes the temporal correlation between the three time series i.e., quarterly recalls of beef, pork and poultry.

Time series diagnostics involved checking for stationarity and autocorrelation using data for the sample period 1994-2008. The null hypothesis of non-stationarity based on the Phillips Perron test is rejected for beef at 5% level and pork and poultry at 1% level. The Ljung-Box Q test results provide significant evidence that there is autocorrelation in the time series. Table 3.2 shows pairwise unconditional correlation between the number of recalls per quarter. The t-test is performed to test the significance of the correlation. Linear unconditional correlation between pork and poultry recalls is the highest at 0.356 and significant. This association could be due to similarities in pork and poultry industry. For example, both pork and poultry are produced in confinement in climate-controlled buildings. Also, both the industries are more vertically integrated and offer more consumer driven value added processed products than beef. The correlation between beef and pork is close at 33% although the correlation is not significant. The correlation between beef and poultry is 25% and significant. One drawback of this time invariant correlation, is that, this will include old information that may be of far less use than recent information. Also, this is a global correlation measure. Local correlations or relationships between a few events in a particular time frame, might not be the same across the entire time frame. This calls for a more time dependent study of the correlation structure.

Correlations between recalls can be examined under two different time scales: (1) Frequency of recalls in each time period (i.e. each quarter) (2) Intra recall event time period (number of days between two recalls). Rolling correlations (Figures 3.2) are used to visualize whether or not the correlations between the frequency of recalls tend to be stable over time. Although using only a few recent observations to calculate these correlations creates more variability, this method is nevertheless useful in obtaining a general idea of the trend in these correlations. The figures show rolling correlations over sliding window widths of 8 quarters. Rolling means and correlation coefficients of 8 quarter window width are computed by starting with the first 8 quarters and then rolling the sample period forward by one quarter at a time. The linear trend line fitted to each graph shows the long run trend in correlation. An analysis of the long term trend emerging from the history of considered temporal correlation shows that there is a negative trend between the correlation of beef and pork recalls, and a positive trend between beef and poultry recalls. Part of the fluctuation in the correlation could be explained by the instability of the mean recalls over the same time periods. During the period 2000 to 2002 when the average number of recalls were large for all three groups the correlations were also large and positive. With the exception of these years, the correlation between poultry and beef recalls is mostly negative while the correlation between poultry and pork recalls is mostly positive. A negative correlation between poultry and beef suggests that when poultry recalls occurred, beef recalls did not. Both the constant and the dynamic correlations between pork and poultry recalls are positive.

Following [64], the stability of the correlation matrix for different time periods is studied. The null hypothesis is that the correlation matrix is constant over two adjacent sub-periods. A brief description of the Jenrich (1970) statistic that tests the equality of two matrices is given separately as Appendix A. Table 3.3 shows the results of the Jennrich test of equality between two correlation matrices calculated over different time periods. The test statistic has an asymptotic chi-square distribution with 28 and 120 degrees of freedom for window sizes 8 and 16 respectively. The null hypothesis of a constant correlation between pork and poultry is rejected 45% of the times for a window size of 4 with a lag of 4 and the average correlation between the two is approximately 0.053. These are correlation between 1994Q1 to 1995Q4 and 1996Q1 to 1997Q4 and so on. The chances of rejecting the null hypothesis of constant correlation decreases with the increase in window size. The greatest number of significant values of the Jennrich test statistic is obtained when the window size equals the lag length. The chances of obtaining a nonconstant correlation over a period of 8 quarters is highest for beef and pork (43.2%), followed by beef and poultry (31.2%) and then pork and poultry (22.7%). Changing structure of the meat and poultry industry could be one reason for the instability of these correlation matrices.

Another way to look at recall correlations is similar to that of bond default correlations.

This is done by calculating the correlation between the number of days survived without recalls. In case of meat recalls, the smallest time difference between successive recalls is one day. There are days on which more than one recall occurs. The correlation between the time lags between two recall events, is calculated using the pairwise binomial correlation used by [44]. Let $I^i_{t_0+T}$ be the indicator function for when a recall for product i occurs T time units after t_0. i takes the value B for beef, P for pork and C for poultry. Then a binary measure of the covariance between beef and pork recalls is given by $E[I^B_{t_0+T} I^P_{t_0+T}] - E[I^B_{t_0+T}]E[I^P_{t_0+T}]$. We choose $t0$ randomly and empirically calculate the hull pair wise binomial correlation statistic given by:

$$\rho = \frac{E[I^B_{t_0+T} = 1, I^P_{t_0+T} = 1] - E[I^B_{t_0+T} = 1]E[I^P_{t_0+T} = 1]}{\sqrt{var[I^B_{t_0+T} = 1]var[I^P_{t_0+T} = 1]}} \tag{3.1}$$

Figure 3.3 shows the correlation between the survival days of the two meat groups varying over the number of days. The figure indicates that correlation path is bi-modal for all three groups. There is zero correlation between beef and pork recall that survived 50 days. For smaller number of days there is positive correlation, reaching almost 8% within the first 25 days. For more than 50 days, the beef and pork recall survival days are likely to be negative reaching a low of -17%. A negative correlation between two species implies that if there is a recall in one species then there possibly would not be any recalls in the other species. Almost the same pattern is observed for pork and poultry with a more enhanced positive correlation, almost 11% within the first 30 days. For poultry and beef survival days the correlations are largely negative with zero correlations observed within the first 10 days and again after almost 60 days.

Some conclusions could be drawn from the correlation analysis of the recall events of beef, pork and poultry. There is enough evidence to show that the null hypotheses of no correlation cannot be completely rejected. Statistical tests for the presence of non-zero correlations in the raw data shows that there is positive and significant correlation between the recalls of pork and poultry products, followed by beef and pork and then beef and poultry. Pork and poultry are more similar in various stages of production, confined, non-ruminant, etc. Also, poultry industry is highly vertically integrated followed closely by the hog industry. It can also be concluded that the conditional correlations between the recalls are likely to be nonconstant over time. Recalls are sources of new information that could potentially change the time structure of future recalls. In the following sections, the effect of ignoring this correlation in consumer demand estimation and price volatility is estimated.

Application in Consumer Demand System

The role of meat product recall as a source of information on perceived quality of meat, and, consequently, the effect on its demand, has been studied extensively. [67] and [56] have included the recall of other meat and poultry products in the estimation of the demand shares. [28] presents a model with correlated learning across products that are marketed under a common brand name. The consumers perceive that the quality of common brand products is correlated and these correlations affect choices. However, to date a formal analysis of the correlation structure between the time of recall of meat and poultry items has not been addressed. In this section, we look at whether the correlation between the timing of food recalls affect consumers perception on any hidden information about food safety issues. First, a Rotterdam model for meat and poultry products is estimated. Second simulations are used to identify the variation in the estimated demand as a result of variation in the correlation structure.

Model Specification

Different authors have studied effects of meat recalls on several different variables. [82], [98], and [92] examined the effect of meat and poultry recalls on firm's stock price, market returns, and societal reactions. They present evidence that the market reacts to food recalls. [67] estimated a Rotterdam model incorporating Food Safety Inspection Service (FSIS) recall information and found a small, but statistically significant decline in meat demand and an increase in demand for non meat goods following meat recalls. The procedure followed by [67] is replicated here with newer data and a different set of exogenous demand shifting variables. All the recalls reported in FSIS data set is included irrespective of their size. The Rotterdam model belongs to the class of differential demand systems and is obtained from a first order approximation to the Marshallian demand functions. The Rotterdam model is of particular interest here because it easily accommodates multiple covariates that may be highly correlated in levels, but not in first differences. The demand equations are generated by defining the total differential equation for each food product, without the need to specify the utility or cost functions. Following [67], the share equation of the Rotterdam model estimated here is given by:

$$w_i dlnx_i = a_{i0} + \sum_{j=1}^{n} a_{ij} D_{ij} + b_i(dln\bar{q}) + \sum_{j=1}^{n} c_{ij}(dlnp_j) + \sum_{k=1}^{K}\sum_{l=1}^{L} \delta_{ikl}(dlnR_{kl}) + v_i \tag{3.2}$$

where w_i is budget share of the i^{th} good (i=1,,4), d is the standard across-period first-difference operator, x_i is per capita consumption of good i, D_j is a quarterly dummy variable included to capture seasonality, p_j is the price of the j^{th} good, $dln\bar{q}$ is a Divisia volume index, R_{kl} represents the kth exogenous shifter with lag length of $l = 0, 1, 2, ...L$, a_{i0}, a_{ij}, b_i and c_{ij} are the intercept and the parameters to be estimated and v_i is a random error term. The intercept or the linear time trend is included for any structural changes that are not captured by the exogenous shifters. Demand restrictions, obtained from economic theory, imposes parameter constraints. The adding up conditions are given by the following set of equations:

$$\sum_{i=1}^{N} c_{ij} = 0 \quad \textstyle\sum_{i=1}^{N} b_i = 1 \tag{3.3}$$

$$\sum_{i=1}^{N} \delta_{ikl} = 0 \quad \textstyle\sum_{i=1}^{N} a_{ij} = 0 \tag{3.4}$$

$$\tag{3.5}$$

The homogeneity and symmetry restrictions are the following:

$$\sum_{j=1}^{N} c_{ij} = 0 \quad c_{ij} = c_{ji} \tag{3.6}$$

Similar to [67], the empirical demand system is specified as a four good demand system that includes beef, pork, poultry, and other consumption goods. This provides flexibility for the meat recall elasticities across beef, pork, and poultry to be negative or positive. One share equation (all other goods) from the demand system is deleted before estimation to avoid singularity in the estimated variance covariance matrix of the error terms. The parameters of this omitted equation are recovered using the adding-up restrictions from economic theory of demand.

Data Description for Demand System Analysis

The quantity of beef, pork and poultry represents quarterly per capita disappearance expressed in retail weights (in pounds). The poultry variable includes both chicken and turkey. The prices are the estimates of quarterly average retail prices (cents per pound). The poultry price is obtained by dividing the poultry expenditure by poultry consumption. But the expenditure variable is available only until 2005 2nd quarter. For the rest of the years, the composite retail price of chicken (a weighted average of whole chicken prices and prices for parts) is used. These price and quantity data are obtained from USDA-ERS. The complete demand system specification includes an aggregate commodities, all other goods. Expenditure on all other goods was calculated using personal consumption expenditure (PCE) per capita less the expenditure on meat and poultry. The PCE data were obtained from the Bureau of Economic Analysis (BEA). The price index for personal consumption expenditure is used as the price of all other goods. The summary statistics of the quarterly data (1994-2008) used to estimate the meat and poultry demand is provided in Table 3.1.

In the Marsh et. al. article, the authors introduced a second measure of product recalls based on popular press covering of meat recalls. They did not find any statistically significant effect of media coverage. In contrast this paper tests whether *media distraction*, that is all the popular news items other than the news on foodborne disease outbreaks, crowd out the information that people get from food product recalls. Theoretically, the idea of media distraction is based on a growing literature in behavioral economics on mental accounting and cognitive limitations ([91]). One way of thinking of this is that we have a limited resource to devote to multiple tasks and our time allocation to media is limited. If the airwaves are dominated by big news event, then either a recall may not get much airing, even if the recall stories are there, or consumers might be too distracted to take into account the information in the recall news items. The quarterly index for media distraction was created using the aggregate news articles published in the World Almanac ([52]). World Almanacs are available from 1993 to 2009 and each almanac has a section on the Chronology of the previous year's event, reported month by month in 3 categories, national, international and general categories. Only the national and the general news articles were aggregated leaving out all the published news on meat/poultry related outbreak leading to recalls . For example, the following news items were not included in the aggregate quarterly index of *media distraction*: 1.ConAgra recalled chicken and turkey pot pies due to 152 cases of salmonella poisoning in 31 states (October, 2007). 2. Ecoli in ground beef from the Topps Meat Company in New Jersey leading to the second largest beef recall in US history. (September, 2007) 3. 5.7 million pounds of potentially contaminated meat were recalled by United Food group (April, May 2007). 4. Ecoli affecting 57 people in 7 states whereby ConAgra recalled 2.8 million pounds of ground beef (August,September 2002). 5. 25 million pounds of beef recalled, by Hudson Foods Company (August 1997). Summary statistics of the *media distraction* index is also shown in Table 3.1. On average there are almost 33 big news items per quarter, ranging from 19 in the third quarter of 2001 to 46 in the second quarter of 2003.

The adding-up constraint implies that only three equations in the system are independent. The procedure followed in this study is to drop the other goods equation, estimate the remaining system, and then calculate the parameters from the omitted equation using the classical restrictions. To obtain estimates of the standard errors of the "deleted" equation, the model was estimated twice: once with the poultry equation deleted, and again with the other good equation deleted.

Results of the consumer demand model

The demand model is estimated using SUR with restrictions on the parameters. The demand equation for other goods is dropped during the estimation to avoid singularity of the error covariance matrix. Some quarters had zero recalls and to perform logarithmic transformations, 1 was added to each FSIS recall. Symmetry, adding up and homogeneity conditions are imposed. The empirical analysis was conducted following [67], where lag length from 0 to 2 are considered for the FSIS recall variables.

The estimated price, expenditure, seasonality and trend coefficients of two Rotterdam models, one with media distraction and the other without media distraction are reported in Table 3.4. The goodness of fit is measured by the adjusted R square which is 83%, 88%, 78% and 99% for beef, pork, poultry and other goods respectively. The matrix of price coefficients for the estimated Rotterdam models is negative semi-definite. Own price coefficients for beef is statistically significant at the 0.05 level. Expenditure coefficients for poultry and other goods are statistically significant at the 0.05 level. All the coefficients of the seasonality and trend variables are statistically significant.

The parameter estimates of the lagged and current recall variables and the media distraction index are reported in the Table 3.5. Current period recall events are negative for beef and pork equations and positive for poultry equation. Statistically significant effects of current period recalls are observed for beef and poultry equations. Lagged values of recalls are not significant for any of the other equations. For the poultry equation, a positive own current period effect is counter-intuitive. A likelihood ratio test was used to compare the two model specifications, one with media distraction index and the other without it. It is found that the media distraction index is not significant. The coefficients of media distraction index is negative for the meat and poultry products and positive for all other goods. The lack of statistical significance for the media distraction index suggests that other news articles do not necessarily crowd out the information obtained from recalls, weakening their impact as a source of information on food safety.

Table 3.6 shows the compensated price and expenditure and current period recall elasticities that are computed at the mean value of the shares from coefficient estimates of the earlier tables. Own-price elasticity coefficients are all negative, indicating the expected inverse relationship between price and quantity demanded. Beef is the most price sensitive at -0.53, followed by pork and poultry at -0.18 and -0.15. [67] also find beef to be the most price elastic at -0.78%. Expenditure elasticities are positive as expected for normal goods, with poultry being more sensitive to expenditure than beef or pork. Expenditure elasticity estimates are 0.25 for beef, 0.13 for pork, 0.95 for poultry and 1.003 for other goods. This result is different from that of [67], who find beef to be the most sensitive to expenditure unlike poultry products. The elasticity of beef demand is negative with respect to recalls of all the species. The elasticity results for poultry demand with respect to recalls are all positive. This counter-intuitive result for poultry suggests the need to look at the long run elasticity of poultry demand with respect to the recalls. One reason these results differ from those of [67] could be that the study period is different. They study recalls from 1982 to 1994, while this paper looks at recalls during the period 1994 to 2008. This analysis included all the recalls irrespective of the size of recalls, similar to the analysis by [67]. More reasonable estimates could be obtained by controlling for the size of the recalls included in the analysis, as consumers are usually aware of massive product recalls as opposed to the smaller recalls that get unnoticed. However these large events are also sparse and would require a different approach to analysis which is a future course this research could take.

Simulation of demand system with correlated recalls

In the following sections, a Monte Carlo experiment is designed to show how the simulated shares of expenditures change when there is a change in the correlation structure between the recalls of the three products, with all the other variables remaining constant at the mean values. The recall events are first modeled as univariate independent Poisson distributions. A Gaussian copula is then used to defined the dependence structure between the Poisson marginal distributions assumed for the recalls. The correlated random draws from Poisson distribution are the correlated recall variables which are plugged into the estimated demand system to obtain the variation in the shares of expenditure. Mean shares do not change as there is no change in the mean values of recalls. However, both the standard deviation and the correlations between the shares changes. Hence, the elasticity of demand with respect to recalls remain unaffected but their standard deviation and hence their confidence interval might change.

Recall events as a Poisson count models

Recall events are a sequence of events that are randomly spaced in time. A simplifying assumption is made for the purposes of simulating recalls, that is recalls for a particular species are assumed to be independent. The recalls of a particular product can be assumed to follow a Poisson process. The arrival rate (λ) of this Poisson process is given by the average number of recalls per unit time. Then the number of recalls (k) in an interval of length t has a Poisson distribution with parameter λt, i.e. $p(k,t) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}$.

Next, a discuss of arrival times is provided to better understand the properties of the underlying process. Let T_j be the time of jth recall. The probability of no recalls in the interval $(t, t+s)$ is given by $e^{-\lambda s}$. Then the event that at least one recall does occur between T_j and T_{j+s} is given by $(1 - e^{-\lambda s})$. The inter arrival times $T_1, T_2 - T_1, T_3 - T_2$... of a Poisson process are i.i.d. with distribution function given by $1 - e^{-\lambda s}$. This is exponentially distributed with density $\lambda e^{-\lambda t}$. Therefore, the inter arrival times of a Poisson process are i.i.d. with an exponential distribution and the converse is also true. Since the exponential density is monotone decreasing, there is a high probability of a short interval and a small probability of a long interval between arrivals ([7]). The recall arrival intensity per quarter, k, is calculated by counting the number of recalls in each quarter from 1994 to 2008. The average recall per quarter for beef is 5.23, for pork it is 1.72 and for poultry it is 3.1.

Dependent Poisson random variables

There are a variety of methods for generating Poisson multivariate random variables. The Trivariate Reduction method proposed by [66], a computationally fast modification of this method presented by [87], a convolution based method by [58] are some of the commonly used methods. The drawbacks of these methods are that they either do not support negative correlation values or involve very complex methodology.

In this paper, a copula based technique is used to generate the correlated Poisson random draws. The copula model for multivariate distributions take into account the effects of the marginal distributions as well as the dependence between them. The copula technique has been introduced in an earlier chapter of this thesis. Copula based models can, in general, be estimated without resorting to numerical integration or simulation. [62] proposed a Gaussian copula function which was used by [95] to analyze dependence in a bivariate count model. A full discussion of the technique is available in [93] parts of which are provided in Appendix B. The correlation matrix used to generate these random numbers take the values of the quarterly correlation coefficient. The simulation strategy assumes a fixed time. Figures 3.4 shows the randomly generated bivariate poisson marginals for the three food groups. The first diagram shows 100 randomly generated numbers of pork and beef recalls per quarter that have a correlation coefficient of -0.09. The linear fit between these randomly generated recall variables show a slight negative relationship.

Results of the simulation of the demand model

As expected, recall correlations directly affect the correlations between shares of expenditure on beef, pork and poultry, as shown in Table 3.7. The table shows pairwise correlation between simulated shares of expenditure for given pairwise correlation between recalls. The expenditure share correlations approach the recall correlations as the simulation sample size is increased, except for the case of perfectly positive correlation between recalls. Table 3.8 shows the mean and the standard deviation of the shares of expenditure. There is very little change in the mean values of shares with respect to changes in recall correlation. However, the standard deviation of the shares increase with the increase in the recall correlation. The percentage increase in the standard deviation is the largest for poultry expenditure share (60%), followed by expenditure share of beef (47%) and then pork (close to zero) if the recall correlation increases from 0% to nearly 100%.

Application in retail price volatility

Commodity price volatility or uncertainty has been widely modeled as the conditional variance in the GARCH framework, originally developed by Engle and later generalized by [10]. Many papers have employed this methodology to explore various issues related to commodity price volatility ([43]). In GARCH models risk is defined as a function of the variance of the price forecast errors conditional on available information. This paper extends the GARCH framework to test the hypothesis that the recalls of the meat and poultry products provide sufficient information to better predict the volatility of the retail prices of these products. A multivariate GARCH would capture the temporal dependence in the second order moments of the prices. It would give a more complete picture of how the price volatilities move together more or less closely over time. However, this paper investigates, more simply, how the second order moments of the recalls of different products affect the volatility of prices and not the changing nature of the relationship between the prices. A univariate GARCH model is a first step towards this analysis. Following [107], an exponential GARCH model is estimated to account for the asymmetry of negative shocks (recalls) that can have a bigger impact on volatility of commodity prices than positive shocks.

Let p_{it} be the price of the meat product i. The first difference of the natural log of price is given by $y_{it} = log(p_{it}/p_{i,t-1})$. The conditional mean of the returns from retail prices are specified as an ARMA(1,1) process to capture any autocorrelation effect present in the market. Price volatility is modeled using exponential GARCH (EGARCH) specification where the logarithmic value of conditional variance is specified as a linear function of the past squared errors, past values of the conditional variance and also include the recall events as the exogenous variable. A EGARCH(1,1) model of [72] for the conditional variance equation is specified as:

$$ln(h_t) = \omega + \beta ln(h_{t-1}) + \gamma\left(\frac{\epsilon_{t-1}}{h_{t-1}^{1/2}}\right) + \alpha\left[\left(\frac{|\epsilon_{t-1}|}{h_{t-1}^{1/2}}\right) - \left(\frac{2}{\pi}\right)^{1/2}\right] + \sum_{i=1}^{3} c^i x_{t-1}^i \tag{3.7}$$

where ω,β,γ and α are coefficients to be estimated.

When ϵ_{t-1} is positive (negative) there is good (bad) news. Bad news can have a larger impact on volatility. The asymmetry is captured by γ. If $\gamma = 0$ there are no asymmetric effects. If γ is positive (negative) high (low) price news generates more volatility. The external regressors x_{t-1}^i are the lagged quarterly recalls of beef, pork and poultry.

Results

The asymmetric volatility models for the quarterly prices of beef, pork and poultry are estimated using R 2.12.0 supplemented with package rgarch. Table 3.9 shows the results of a standard likelihood ratio test, where under null hypothesis, the three coefficients of the recall events, for each equation, are constrained to zero, (the standard EGARCH(1,1) without the exogenous variables.) The alternative hypothesis is the unconstrained GARCH model which includes the recall variables. The appropriate statistic is twice the difference of the maximized values of the log likelihood functions for the unconstrained and constrained models, respectively, which will have a chi-square distribution with 3 degrees of freedom under the null hypothesis. The alternative hypothesis that the recall events explain a part of the nonconstant conditional variance could be rejected for the beef prices. The recalls variables are significant in explaining the volatility of pork and poultry prices.

Maximum Likelihood estimates of the EGARCH(1,1) model without the recall variables are shown in Table 3.10. There is evidence of nonconstant variance since the GARCH effects is high and significant for each of the food groups. The sum of the ARCH coefficient and γ gives the effect of high price news on conditional variance. It is 0.26 for beef, 0.35 for pork and 0.24 for poultry. That is, an unexpected price increase with $\epsilon_{t-1} > 0$ increases the volatility for pork by 35%, followed by beef and poultry. The effect of low price news is given by the difference between the ARCH term and γ, i.e. 0.25 for beef, -0.40 for pork and -0.31 for poultry. An unexpected price decrease with $\epsilon_{t-1} < 0$ decreases the volatility of pork by 40%, poultry by 31% and increases the volatility of beef by 25%. For pork and poultry prices, the negative shocks have a bigger impact on volatility than the positive shocks.

The estimated effects of recall variables on the volatility of prices are shown in Table 3.11 which presents the results of the EGARCH(1,1) model with recalls. In both the models the GARCH effects are large and positive although there is a marked decline in the case of pork. Beef recalls increase the volatility of beef and poultry prices but have a significant decreasing effect on the volatility of pork price. Pork recalls have a significant negative effect on poultry prices. Poultry recalls decrease the volatility of poultry and beef prices and these effects are significant. The asymmetry terms in this model are negative for beef and pork prices and positive for poultry prices. An unexpected price increase ($\epsilon_{t-1} > 0$) will decrease the volatility of pork by almost 8%, beef by 30% and poultry by 19%. A unexpected price decrease ($\epsilon_{t-1} < 0$) will increase the volatility of pork by more than 100% and that of beef by 80% and decrease the volatility of poultry by 4%.

Conclusion

This study investigates the presence of correlation between recalls of closely related food products to examine the claim that the recalls appear in clusters over time. Both the static and dynamic correlations between the FSIS recall events of beef, pork and poultry are estimated. There is evidence that the recalls are correlated across products. The empirical findings show the significant and positive pairwise correlation between pork and poultry recalls, and pork and beef recalls highlighting the similarities in these industries. Moreover, this correlation is time dependent and decreases with the increase in the number of days between two consecutive recalls.

The paper then proceeds to examine the effect of these recall correlations on the estimation of the shares of expenditure in the meat/poultry consumer demand model. The parameter estimates and resulting elasticity coefficients are largely consistent with classical demand theory. There are two ways in which food recall events might affect a consumer's demand. First the recall event itself might lead the consumer to switch within product brand or the cross product loyalty. This effect is well established in the existing literature, where the own elasticity of recalls on demand are negative. This paper investigated whether people are distracted by other media headlines when responding to recalls. It was found that the media distraction index was not significant. Thus there is no evidence of crowding out of the information contained in food recalls due to other news headlines dominating the airwaves.

A simple simulation is provided to understand the effect of current period cross correlation on the estimated shares of expenditure. The results of the simulation are consistent with expected values of the mean, standard deviation and correlations of the shares of expenditure.

It is found that the variance of the expenditure shares increases with the increase in the correlation between the recalls. These findings imply an increase in uncertainty during periods of highly clustered recalls.

Asymmetric GARCH models of volatility of the commodity prices as a function of recalls are used to look for evidence of additional risk. Some of the volatility in the quarterly retail prices can be explained by the recalls. This is especially true for pork prices. For all beef and pork, the effect of bad news on the volatility of prices is more pronounced when recalls are included in estimation. The asymmetric effect on poultry prices are not amplifying when recalls are included. One explanation could be that the poultry markets are matured as suggested by [107] who put forth the idea that mature food markets have constant variance. It can be concluded from this analysis that recalls send food safety signals that affect consumer demand decisions and volatility of commodity prices.

Table 3.1 Summary statistics of quarterly data (1994-2008)

Variable	Mean	Standard Deviation	Minimum	Maximum
Beef consumption (lbs/capita)	16.5	0.58	15	17.5
Pork consumption (lbs/capita)	12.7	0.68	11.3	14.3
Poultry consumption (lbs/capita)	24.2	1.68	21	27.1
Retail beef price (cents/lb)*	340.2	58.57	273.5	445.9
Retail pork price (cents/lb)*	257.4	26.88	201.2	300.8
Retail poultry price (cents/lb)*	158.6	9.81	140.6	177.3
Beef expenditure share (%)	31	1.51	27.4	34.2
Pork expenditure share (%)	23.7	1.17	21.5	26.5
Poultry expenditure share (%)	45.2	1.82	41.8	48.1
Beef Recalls per quarter	5.2	3.82	17	314
Pork Recalls per quarter	2.15	1.77	7	103
Poultry Recalls per quarter	3.13	2.05	9	186
Media distraction index	33.83	5.97	19	46

Table 3.2 Correlation between quarterly recalls by food groups

Variable	by Variable	Correlation	Count	Lower 95%	Upper 95%	Signif Prob
Pork	Beef	0.3292	60	0.0822	0.5382	0.0102*
Poultry	Beef	0.2566	60	0.0029	0.4793	0.0478*
Poultry	Pork	0.3560	60	0.1122	0.5594	0.0052*

One asterisks indicate statistical significance at 10% levels

Table 3.3 Test of the equality of correlation matrix over time

Pairs	Window Size	X^*	Average Correlation
pork&poultry	4	25	0.053
	8	22.72	0.0099
	16	10.71	-0.39
poultry&beef	4	36.53	-0.087
	8	31.81	-0.096
	16	28.57	-0.34
beef&pork	4	34.61	0.03
	8	43.18	0.23
	16	32.14	-0.098

*The null hypothesis of a constant correlation matrix is rejected at the 15% confidence level in X% of the times

Table 3.4 Price, expenditure, seasonality and trend coefficient estimates of the two Rotterdam demand models (t statistics reported below)

	Demand Equation with media distraction				Demand equation without media distraction			
Dependent Variable	Beef	Pork	Poultry	Other goods	Beef	Pork	Poultry	Other goods
Beef price	-0.00119				-0.00120			
	-5.50				-10.78			
Pork price	0.00004	-0.00024			0.00004	-0.00024		
	0.21	-0.74			0.19	-0.74		
Poultry price	0.00010	0.00011	-0.00022		0.00009	0.00011	-0.00022	
	0.54	0.57	-0.95		0.59	0.57	-0.95278	
Other price	0.00105	0.00009	0.00001	-0.00116	0.00107	0.00010	0.00002	-0.00119
	0.001	0.238	0.032	-2.276	3.37	0.25	0.048	-2.3261533
Expenditure	0.00033	0.00011	0.00126	0.99829	0.00056	0.00018	0.00139	0.99787
	0.59	0.18	2.05	1096.47	0.94	0.30	2.35	1133.0959
Quarter 1 dummy	0.00011	-0.00023	-0.00011	0.00023	0.00010	-0.00023	-0.00012	0.000245
	6.75	-13.38	-6.55	9.18	6.71	-14.34	-7.18	9.967891
Quarter 2 dummy	0.00026	-0.00016	0.00003	-0.00013	0.00025	-0.00016	0.00003	-0.000118
	15.75	-9.13	1.85	-5.33	16.11	-9.66	1.70	-5.030989
Quarter 3 dummy	0.00010	-0.00009	-0.00001	-0.0000029	0.00010	-0.00009	-0.00001	-0.000004
	6.30	-4.50	-0.87	-0.11	6.34	-4.50	1.70362	-0.1526093
Intercept	-0.00012	0.00012	0.00002	-0.00002	-0.00011	0.00012	0.000023	-0.000028
	-10.83	10.13	1.93	-1.39	-10.78	10.33	2.09	-1.6755882
Adjusted R square	0.831	0.883	0.785	0.990	0.83	0.88	0.79	0.9900

Table 3.5 Recall and media distraction coefficient estimates of the two models (t statistics reported below)

	Demand Equation with media distraction				Demand equation without media distraction			
Dependent Variable	Beef	Pork	Poultry	Other goods	Beef	Pork	Poultry	Other goods
Beef Recall (L=0)	-0.00002	-0.000005	0.00002	0.0000037	-0.000023	-0.000006	0.000023	0.000005
	-1.84	-0.45	1.95	0.21	-1.89	-0.46	1.93	0.27
Pork Recall (L=0)	-0.00001	0.000000	0.00003	-0.00002	-0.000010	0.000001	0.000030	-0.000021
	-1.38	-0.03	3.17	-1.17	-1.10	0.07	3.47	-1.63
Poultry Recall (L=0)	-0.00001	0.00002	0.00004	-0.00004	-0.000013	0.000016	0.000035	-0.000038
	-1.24	1.71	3.71	-2.81	-1.37	1.68	3.65	-2.65
Beef Recall (L=1)	0.000003	0.000002	0.00000	-0.0000004	0.000004	0.000002	-0.000004	-0.000003
	0.25	0.14	-0.36	-0.02	0.35	0.17	-0.30	-0.20
Pork Recall (L=1)	0.00001	0.00001	-0.00002	0.0000014	0.000012	0.000005	-0.000022	0.000005
	1.51	0.62	-2.33	0.10	1.33	0.57	-2.49	0.37
Poultry Recall (L=1)	0.00001	-0.000001	-0.00002	-0.00001	0.000010	-0.000001	-0.000017	0.000008
	0.90	-0.14	-1.90	-0.71	1.04	-0.10	-1.84	0.60
Beef Recall (L=2)	0.000001	0.00001	0.00001	0.00001	-0.000001	0.000008	0.000004	-0.000012
	0.10	0.90	0.53	0.75	-0.08	0.85	0.44	-0.821
Pork Recall (L=2)	-0.00001	-0.00001	0.00001	-0.00002	-0.000007	-0.000006	0.000007	0.000005
	-0.88	-0.69	0.72	-1.12	-0.75	-0.65	0.80	0.40
Poultry Recall (L=2)	-0.000020	0.0000008	0.00000	0.00001	-0.000022	0.0000005	0.0000001	0.000021
	-2.32	0.09	0.09	0.59	-2.48	0.05	0.01	1.61
Media Distraction	-0.000031	-0.00001	-0.00002	0.00002				
	-1.37	-0.42	-0.77	0.57				

Table 3.6 Compensated price and income elasticities

Compensated Price and Income Elasticity				
	Quantity of:			
With respect to:	Beef	Pork	Poultry	Other goods
Beef Price	-0.529	0.026	0.064	0.001
Pork Price	0.015	-0.179	0.072	0.000
Poultry Price	0.041	0.08	-0.147	0.000
Other goods price	0.472	0.073	0.011	-0.001
Expenditure	0.246	0.134	0.946	1.003
Current period FSIS recall elasticities				
	Quantity of:			
With respect to:	Beef	Pork	Poultry	Other goods
Beef Recalls	-0.0099	-0.0041	0.0158	0.000005
Pork Recalls	-0.0043	0.0005	0.0206	-0.000021
Poultry Recalls	-0.0058	0.012	0.0237	-0.000038

Elasticities are calculated at the mean values of the explanatory variables.

Table 3.7 Correlation between simulated shares of beef, pork and poultry expenditure

Percentage correlation between simulated shares of expenditure				
Recall correlation	Sample size	Beef-Pork	Pork-Poultry	Poultry-Beef
From data*	100	24.60	17.31	22.35
	1000	28.78	25.14	22.55
	10000	31.41	30.96	30.03
0	100	-11.89	-9.23	-8.15
	1000	3.26	-1.83	0.69
	10000	-0.06	-0.34	1.05
0.5	100	71.84	65.50	61.19
	1000	57.14	52.94	50.36
	10000	44.97	44.92	45.09
0.98	100	92.55	87.51	92.29
	1000	91.27	91.16	89.89
	10000	90.74	90.79	90.63

*Correlation obtained from data: Beef-Pork 33%, Pork-Poultry 35% and Poultry-Beef 25%

Table 3.8 Mean and standard deviation of the simulated shares of expenditure

		Beef share		Pork share		Poultry share	
Correlation	Sample size	Mean	SD	Mean	SD	Mean	SD
From data*	100	0.0050755	0.03149963	0.001249	0.0092264	0.001011	0.015076
	1000	0.0007912	0.0259982	0.001288	0.010186	0.0010057	0.013269
	10000	0.0008128	0.02515622	0.001309	0.0102697	0.0010073	0.012678
0	100	0.0008014	0.02387997	0.00111	0.0116603	0.000988	0.00998
	1000	0.000786	0.0218515	0.001321	0.0111291	0.001002	0.010126
	10000	0.00812	0.0214082	0.001309	0.0104769	0.001008	0.010174
0.98	100	0.0006721	0.04300416	0.001114	0.0113425	0.0009978	0.021696
	1000	0.000791	0.03328504	0.001286	0.0100606	0.0009992	0.01735
	10000	0.0008144	0.03192701	0.001309	0.0097829	0.0010071	0.016558

*Correlation obtained from data: Beef-Pork 33%, Pork-Poultry 35% and Poultry-Beef 25%

Table 3.9 Results of the Likelihood Ratio Test

	Value of Log Likelihood Function			
Variable	Under H0 EGARCH(1,1)	Under H1 EGARCH(1,1) with recall variables	Value of test statistic	Result of test
Beef retail price	250.8692	249.8267	2.085	Cannot reject H0
Pork retail price	229.1332	236.0047	-13.743	Reject H0
Poultry retail price	197.0335	183.5644	26.9382	Reject H0

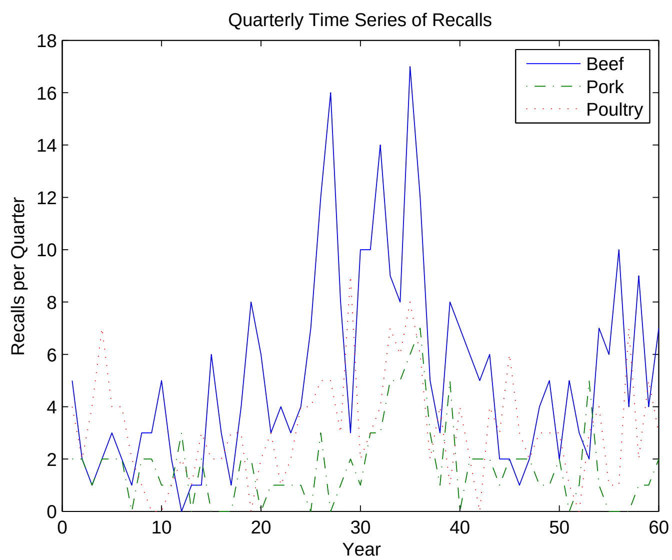
Table 3.10 Maximum likelihood estimates of the Exponential GARCH (1,1) Model

	Beef Retail Price		Pork Retail Price		Poultry Retail Price	
	Coefficients	t value	Coefficients	t value	Coefficients	t value
Conditional mean equation						
Constant	0.006287	2.86742	NA	NA	NA	NA
AR term	-0.387843	-3.03812	0.57375	1.46769	-0.04224	-0.59211
MA term	0.47484	3.2284	-0.31868	-0.61539	NA	NA
Conditional variance equation						
Constant	-0.918473	-1.14841	-0.87718	-0.14433	-0.25902	-1.75292
ARCH	0.253356	1.37418	-0.02607	-0.10207	-0.03137	-0.2592
Asymmetry term	0.003109	0.13387	0.37844	0.28446	0.275552	2.23856
GARCH	0.878158	8.20143	0.87602	1.02464	0.950948	41.05884

Table 3.11 Maximum likelihood estimates of the Exponential GARCH (1,1) Model with recalls

	Beef Retail Price		Pork Retail Price		Poultry Retail Price	
	Coefficients	t value	Coefficients	t value	Coefficients	t value
Conditional variance equation						
Constant	-2.256	-1.630	-2.875	-2.959	-0.504	-3.792
ARCH	0.558	2.919	0.549	2.737	0.076	0.803
Asymmetry term	-0.258	-1.197	-0.470	-0.978	0.118	0.787
GARCH	0.728	4.090	0.628	4.609	0.940	47.425
Beef Recall (lag 1)	0.028	0.866	-0.101	-3.110	0.116	2.501
Pork Recall (lag1)	0.005	0.315	-0.009	-0.484	-0.041	-2.363
Poultry Recall (lag 1)	-0.055	-1.834	0.005	0.251	-0.066	-2.347

Figure 3.1 Total number of FSIS product recalls per quarter by meat category



Data First Quarter in 1994 - Last quarter in 2008.

Figure 3.2 Rolling correlations for pairwise quarterly recalls, window width = 8

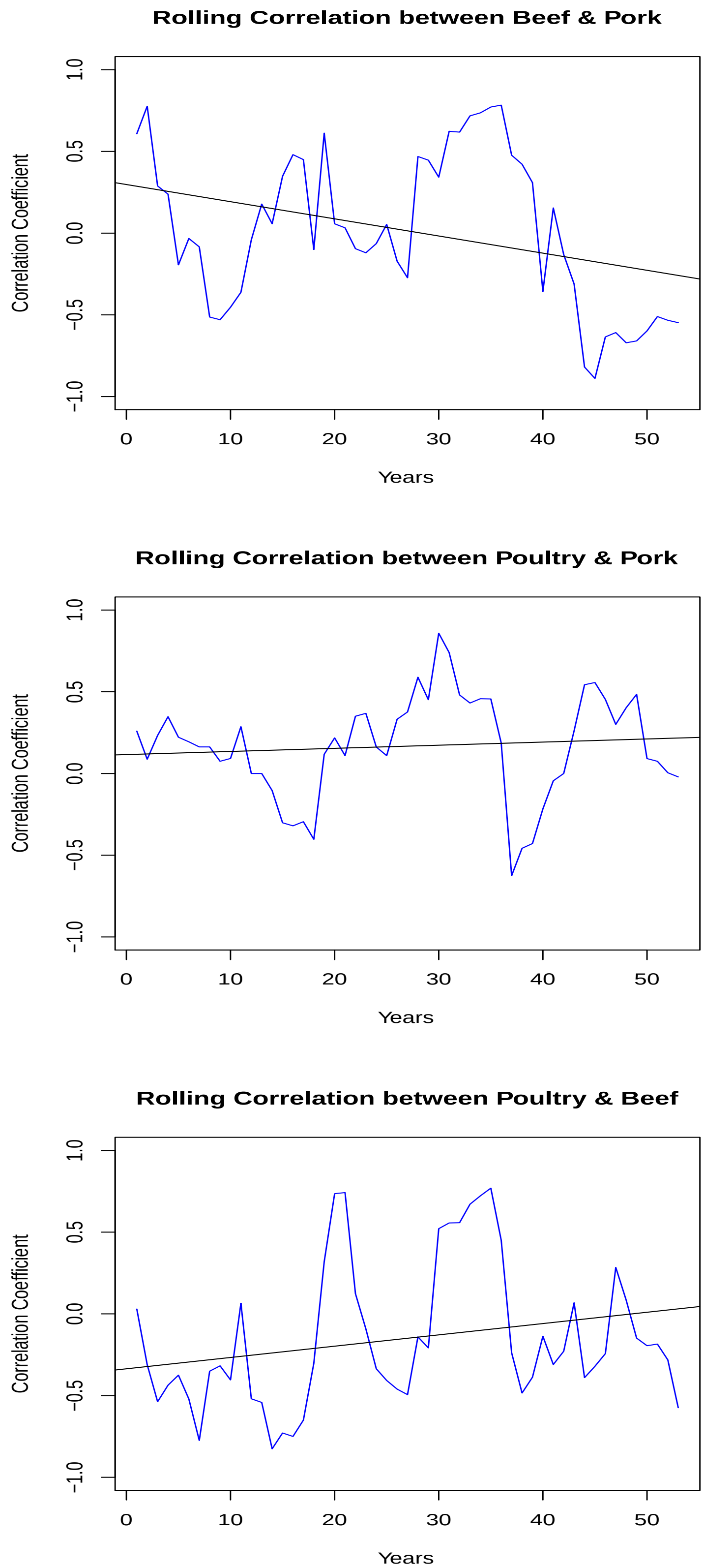


Figure 3.3 Pairwise correlation between the time lags of recall events

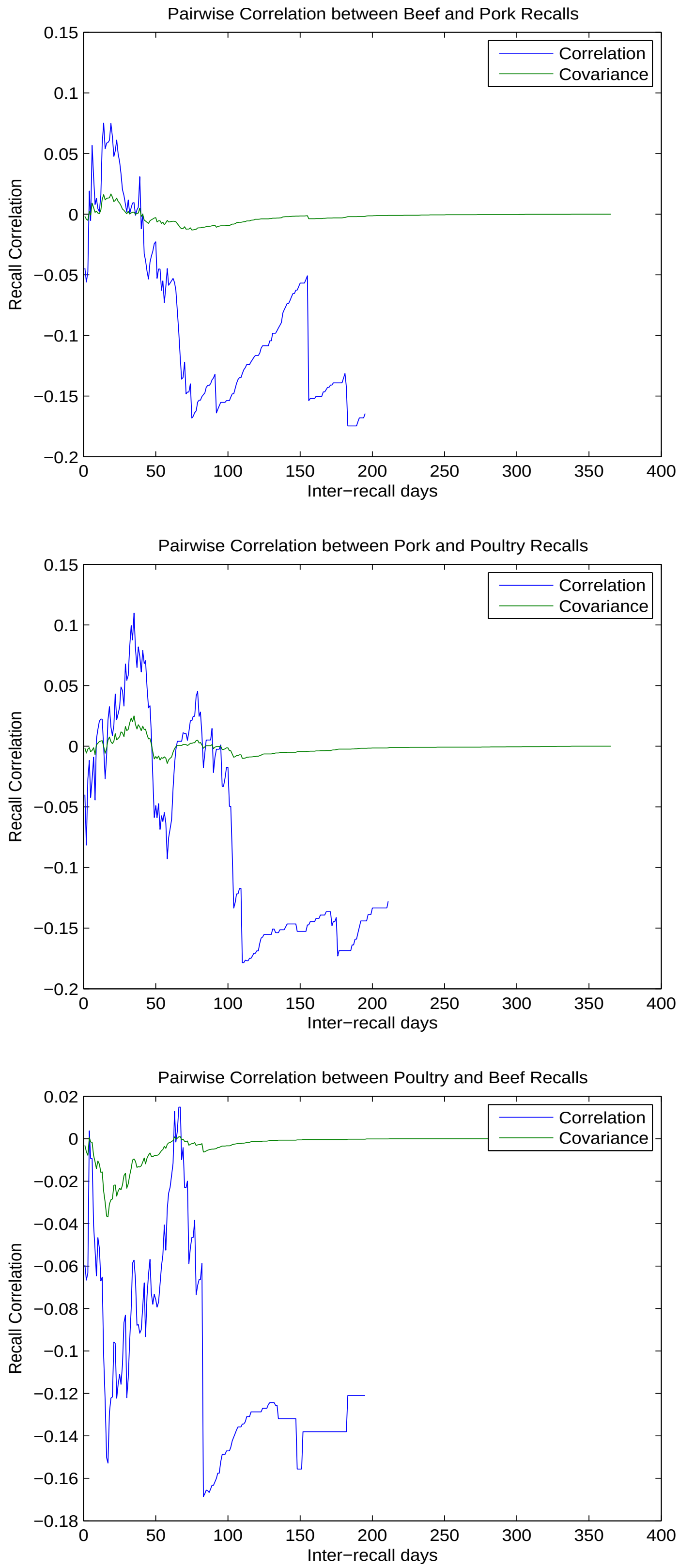
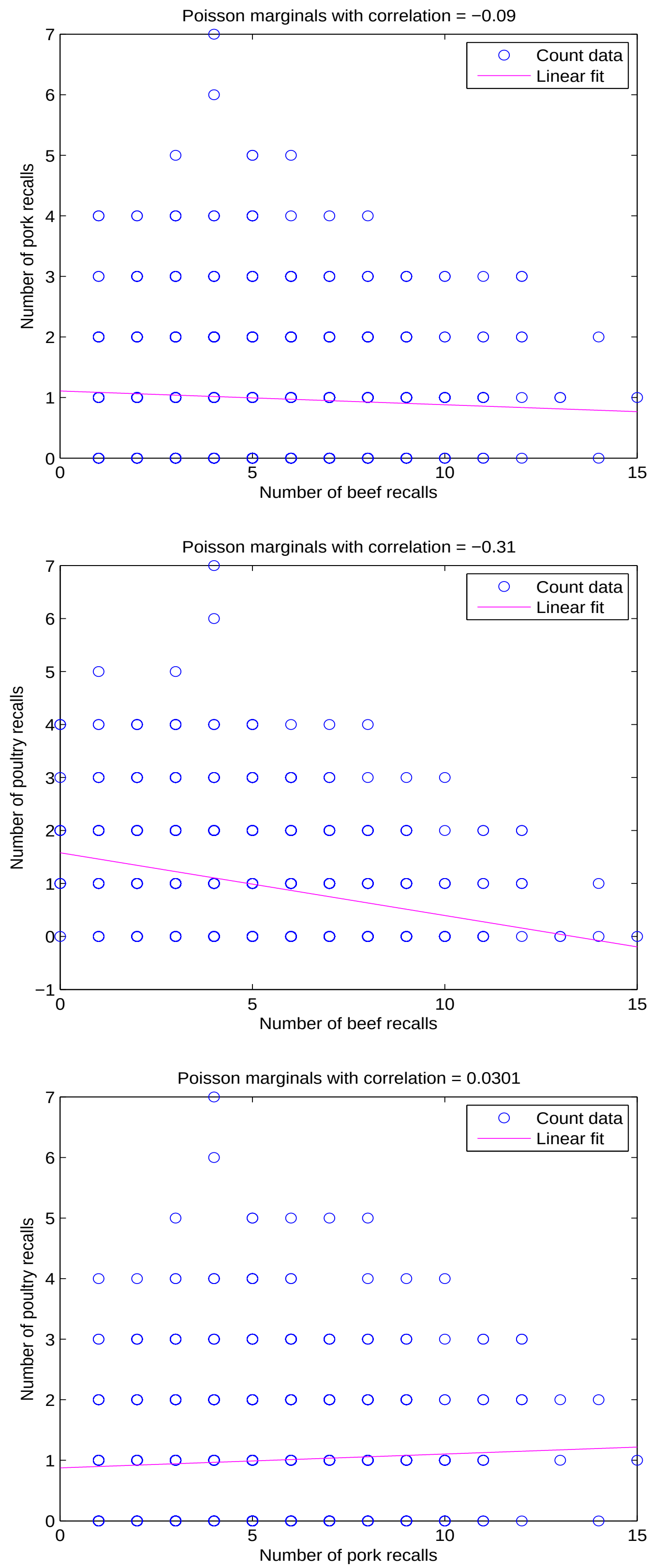


Figure 3.4 Bivariate simulated data generated with Poisson marginals and empirical correlation coefficient



APPENDIX A. Equality of two correlation matrices (Jenrich (1970) test)

The test statistic suggested by [48] to test the equality of two given correlation matrices is shown below. The correlation matrices are computed from different sample sizes. The null hypothesis of the test holds if the two correlation matrices are equal. This test has a χ^2 distribution with the degrees of freedom $p(p-1)/2$, where p is the dimension of the correlation matrix. The test statistic is given by the following equation:

$$\chi^2 = \frac{1}{2}tr(Z^2) - diag'(Z)S^{-1}diag'(Z) \tag{A.1}$$

where,

$$Z = c^{0.5}\bar{R}^{-1}(R_1 - R_2) \tag{A.2}$$

$$\bar{R} \equiv (\bar{r}_{ij}) = \frac{1}{(n_1 + n_2)}(n_1 R_1 + n_2 R_2) \tag{A.3}$$

$$S = (\delta_{i,j} + \bar{r}_{i,j}\bar{r}^{i,j}) \tag{A.4}$$

$$c = \frac{n_1 n_2}{(n_1 + n_2)} \tag{A.5}$$

The term tr in the above expression stands for the trace of the matrix, where diag implies diagonal elements of the matrix. R_1 and R_2 are sample correlation matrices of two successive samples of size n_1 and n_2, respectively, and δ_{ij} is the Kronecker delta (equalling 1 when $i = 1$ and 0 otherwise). The elements $(\bar{r}^{i,j})$ are from the inverse of the matrix $\bar{R}$.

APPENDIX B. Simulation of Random Variables

The following algorithm generates random variables u_1 and u_2 from the Gaussian copula $C(u_1, u_2, \theta)$.

1. Generate two independent distribution $N(0, 1)$ variables v_1 and v_2.
2. Set $y_1 = v_1$.
3. Set $y_2 = v_1\theta + v_2\sqrt{1-\theta^2}$.
4. Set $u_i = \Phi(y_i)$, i = 1,2. Φ is cumulative distribution function of the standard normal distribution.

Then the pair (u_1, u_2) are uniformly distributed variables drawn from the Gaussian copula.

Simulation of Bivariate Poisson.

Technique of [21] to simulate two correlated discrete count variables.

1. Draw correlated uniform random variables (u_1, u_2) from a particular copula using Gaussian method.
2. Set the Poisson mean = μ_1 s.t. $Pr(Y_1 = 0) = e^{-\mu_1}$.
3. Set $Y_1 = 0$, $P_0 = e^{-\mu_1}$, $S = P_0$.
4. If $u_1 < S$, then Y_1 remains 0.
5. If $u_1 > S$, then proceed sequentially as follows: While $u_1 > S$, replace (i) Y_1 with $Y_1 + 1$. (ii) P_0 with $\frac{\mu_1 P_0}{Y_1}$. (iii) S with $S + P_0$. This process continues until $u_1 < S$.

These steps produce a simulated variable Y_1 that is Poisson distributed with mean μ_1. To obtain draws of the second poisson variable Y_2, replace u_1 and μ_1 with u_2 and μ_2 and repeat the steps. Then the pair (Y_1, Y_2) are jointly distributed poisson variables with means μ_1 and μ_2.

BIBLIOGRAPHY

[1] K.E. Arbuckle and J.A. Downing. The influence of watershed land use on lake N: P in a predominantly agricultural landscape. *Limnology and Oceanography*, 46(4):970–975, 2001.

[2] B. A. Babcock, P. W. Gassman, M. Jha, and C. L. Kling. *Adoption subsidies and environmental impacts of alternative energy crops.* Center for Agricultural and Rural Development, Iowa State University, Ames, Iowa., 2007.

[3] B. A. Babcock, P. G. Lakshminarayan, J. Wu, and D. Zilberman. Targeting Tools for the Purchase of Environmental Amenities. *Land Economics*, 73(3):325–339, 1997.

[4] Bruce A. Babcock and David A. Hennessy. Input demand under yield and revenue insurance. *American Journal of Agricultural Economics*, 78(2):416–427, 1996.

[5] Bruce A. Babcock and David A. Hennessy. Getting more acres from the corn belt. *Iowa Ag Review*, 12(4):6–7, 2006.

[6] R. W. Bachmann and R. A. Bachmann. *A classification of Iowa's lakes for restoration.* Iowa State University, 1994.

[7] M. Bagnoli and T. Bergstrom. Log-concave probability and its applications. *Economic theory*, 26(2):445–469, 2005.

[8] J.C. Beghin. (fapri) 2006 U.S. and world agricultural outlook. Center for Agricultural and Rural Development (CARD) Publications 06-fsr1, Center for Agricultural and Rural Development (CARD) at Iowa State University, January 2006. available at http://ideas.repec.org/p/ias/cpaper/06-fsr1.html.

[9] A. Bhaskar and J. Beghin. How coupled are decoupled farm payments? A review of the evidence. *Journal of Agricultural and Resource Economics*, 34(1):130–153, 2009.

[10] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.

[11] T. Bollerslev, R. F. Engle, and J. M. Wooldridge. A capital asset pricing model with time-varying covariances. *The Journal of Political Economy*, 96(1), 1988.

[12] S. T. Buccola. Testing for nonnormality in farm net returns. *American Journal of Agricultural Economics*, 68:334–343, 1986.

[13] M. Burton and T. Young. The impact of BSE on the demand for beef and other meats in Great Britain. *Applied Economics*, 28(6):687–693, 1996.

[14] C. Chatfield. Model uncertainty, data mining and statistical inference. *Journal of the Royal Statistical Society*, Series A(158):419–466, 1995.

[15] N. H. Chau and H. de Gorter. Disentangling the consequences of direct payment schemes in agriculture on fixed costs, exit decisions, and output. *American Journal of Agricultural Economics*, 87(5):1174–1181, 2005.

[16] U. Cherubini, E. Luciano, and W. Vecchiato. *Copula Methods in Finance*. John Wiley and Sons, Hoboken NJ, 2004.

[17] WR Clark, B. Falk, RA Schmitz, and B.A. Babcock. Projecting the Wildlife and Economics Impacts of United States Agricultural Policy Using Landscape-Level Analyses. *Staff General Research Papers*, 2001.

[18] M. Creel and J. Loomis. Theoretical and empirical advantages of truncated count data estimators for analysis of deer hunting in california. *American Journal of Agricultural Economics*, 72, 1990.

[19] R. H. Day. Probability distribution of field crop yields. *Journal of Farm Economics*, 47:713–741, 1965.

[20] H. De Gorter. Export Subsidies: Agricultural Policy Reform and Developing Countries. *A Window into the Issues*, page 109, 2006.

[21] L. Devroye and L. Devroye. *Non-uniform random variate generation*. Citeseer, 1986.

[22] M. D. Duffy. Estimated costs of crop production in iowa. Technical report, Iowa State University, Extension Publication, 2007.

[23] B. Efron. *The jackknife, the Bootstrap, and other Resampling Plans*. Philadelphia: Society for Industrial and Applied Mathematics, 1982.

[24] B. Efron and G. Gong. A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*, 37(1):36–48, 1983.

[25] B. Efron and R. Tibshirani. *An Introduction to the Bootstrap*. London: Chappman and Hall, 1993.

[26] P. Embrechts, A. McNeil, and D. Straumann. Correlation and Dependence in Risk Management: Properties and Pitfalls. *Risk Management: Value at Risk and Beyond*, pages 176–223, 2002.

[27] R. F. Engle and V. K. Ng. Measuring and testing the impact of news on volatility. *The Journal of Finance*, 48(5):1749–1778, 1993.

[28] T. Erdem. An experimental analysis of umbrella branding. *Journal of Marketing Research*, 35:339–351, 1996.

[29] J. Faraway. Data splitting strategies for reducing the effect of model selection on inference. *Computing Science and Statistics*, 30, 1998.

[30] H. Feng, C. L. Kling, L. A. Kurkalova, S. Secchi, and P. W. Gassman. The Conservation Reserve Program in the Presence of a Working Land Alternative: Implications for Environmental Quality, Program Participation, and Income Transfer. *American Journal of Agricultural Economics*, 87(5):1231–1238, 2005.

[31] P. Fousekis and G. Karagiannis. Measuring and Decomposing Productivity Growth in Oligopsonistic Industries: An Application to US Meat Packing Industry. *RISEC: International Review of Economics and Business*, 52(5):223–44, 2005.

[32] Gardner. Futures prices in supply analysis. *American Journal of Agricultural Economics*, 58:81–84, 1976.

[33] G. Gong. Cross-validation, the jackknife, and the bootstrap: Excess error estimation in forward logistic regression. *Journal of American Statistical Association*, 81(393):108–113, 1986.

[34] B. K. Goodwin, A. M. Featherstone, and K. Zeuli. Producer experience, learning by doing, and yield performance. *American Journal of Agricultural Economics*, 84(3):660–678, 2002.

[35] B. K. Goodwin and A. K. Mishra. An Empirical Evaluation of Yield Performance and Cross-Crop Yield Correlation. In *2002 Annual Meeting, July 28-31, 2002, Long Beach, California*. Western Agricultural Economics Association, 2002.

[36] A. W. Gray, M. D. Boehlje, B. A. Gloy, and S. P. Slinsky. How U.S. farm programs and crop revenue insurance affect returns to farm land. *Review of Agricultural Economics*, 26(2):238–253, 2004.

[37] G. Green, D. Sunding, D. Zilberman, and D. Parker. Explaining irrigation technology choices: A microparameter approach. *American Journal of Agricultural Economics*, 78(4):1064–1072, 1996.

[38] C. N. Haas. On modeling correlated random variables in risk assessment. *Risk Analysis*, 19(6):1205–1214, 1999.

[39] C. E. Hart, D. J. Hayes, and B. A. Babcock. Insuring eggs in baskets: Should the government insure individual risks? *Canadian Journal of Agricultural Economics*, 54:121–137, 2006.

[40] D. A. Hennessy, B. A. Babcock, and D. J. Hayes. Budgetary and producer welfare effects of revenue insurance. *American Journal of Agricultural Economics*, 79(3):1024–1034, 1997.

[41] D. A. Hennessy, J. Roosen, and J. A. Miranowski. Leadership and the provision of safe food. *American Journal of Agricultural Economics*, 83(4):862–874, 2001.

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

[43] M. T. Holt and S. V. Aradhyula. Price risk in supply equations: An application of GARCH time-series models to the US broiler market. *Southern Economic Journal*, pages 230–242, 1990.

[44] J. C. Hull. *Options, futures and other derivatives.* Pearson Prentice Hall, 2009.

[45] IDNR. Trends in iowa wildlife populations and harvest - 2004. Technical report, Iowa Department of Natural Resources - Wildlife Bureau, Des Moines, Iowa., 2005.

[46] R. L. Iman and W. J. Conover. Distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics*, B11(3):311–334, 1982.

[47] M. Jansen, J. Thissen, and J. Withagen. USAGE Manual uncertainty and sensitivity analysis in a GenStat environment. Technical report, Biometris, Wageningen UR, 2005.

[48] R.I. Jenrich. An asymptotic Chi Square test for the equality of two correlation matrices. *Journal of the American Statistical Association*, 65:904–912.

[49] Y. Jeon and J. A. Herriges. Convergent Validity of Contingent Behavior Responses in Models of Recreation Demand. *Environmental and Resource Economics*, 45(2):223–250, 2010.

[50] M. E. Johnson and A. Tenenbein. A bivariate distribution family with specified marginals. *Journal of the American Statistical Association*, 76(1):198–201, 1981.

[51] T. Josling. Domestic farm policies and the WTO negotiations on domestic support. Capri (Italy), June 23-26 2003. International Conference on Agricultural Policy Reform and the WTO: Where are we heading?

[52] C. A. Joyce, S. Janssen, and M. L. Liu. *The world almanac and book of facts, 2009.* World Almanac Books, 2009.

[53] R.E. Just and G.C. Rausser. Commodity price forecasting with large-scale econometrics models and the futures market. *American Journal of Agricultural Economics*, 63:197–207, 1981.

[54] Richard E. Just and Quinn Weninger. Are crop yields normally distribued ? *American Journal of Agricultural Economics*, 81(2):287–304, 1999.

[55] M. Khanna, M. Isik, and D. Zilberman. Cost-effectiveness of alternative green payment policies for conservation technology adoption with heterogeneous land quality. *Agricultural economics*, 27(2):157–174, 2002.

[56] H. W. Kinnucan, H. Xiao, C. J. Hsia, and J. D. Jackson. Effects of health information and generic advertising on US meat demand. *American Journal of Agricultural Economics*, 79(1):13–23, 1997.

[57] C. L. Kling, H. Feng, P. W. Gassman, M. Jha, L. A. Kurkalova, and S. Secchi. Natural Resources at Risk: Water Quality and the Dead Zone in the Gulf of Mexico. In *AERE (Association of Environmental and Resource Economists) Workshop, Jackson, WY*, 2005.

[58] F. Krummenauer. Limit theorems for multivariate discrete distributions. *Metrika*, 47(1):47–69, 1998.

[59] L. Kurkalova, C. Kling, and J. Zhao. Green subsidies in agriculture: Estimating the adoption costs of conservation tillage from observed behavior. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie*, 54(2):247–267, 2006.

[60] L. Kurkalova, C. L. Kling, and J. Zhao. Multiple Benefits of Carbon-Friendly Agricultural Practices: Empirical Assessment of Conservation Tillage. *Environmental Management*, 33(4):519–527, 2004.

[61] E. Leamer. Let's take the con out of econometrics. *American Economic Review*, 73(1):31–43, 1983.

[62] L. F. Lee. Generalized econometric models with selectivity. *Econometrica: Journal of the Econometric Society*, 51(2):507–512, 1983.

[63] F. Lindskog. Linear correlation estimation. *Risklab, ETH Zurich*, 2000.

[64] F. Longin and B. Solnik. Is the correlation in international equity returns constant: 1960–1990? *Journal of international money and finance*, 14(1):3–26, 1995.

[65] J. L. Lusk and T. C. Schroeder. Effects of meat recalls on futures market prices. *Agricultural and Resource Economics Review*, 31(1):47–58, 2002.

[66] K.V. Mardia. *Families of Bivariate Disributions*. Charles Griffin and Sons,, 1970.

[67] T. L. Marsh, T. C. Schroeder, and J. Mintert. Impacts of meat product recalls on consumer demand in the USA. *Applied Economics*, 36(9):897–909, 2004.

[68] S. Marwah. Land preservation decisions: Theoretical and empirical analysis. PhD. Dissertation, 2004.

[69] R. B. Nelsen. An introduction to copulas, volume 139 of Lecture Notes in Statistics. *Springer-Verlag, New York*, 1:18, 1999.

[70] R.B. Nelsen. *An Introduction to Copulas*. Lecture notes in Statistics. Springer-Verlag New York, Inc., 1999.

[71] C.H. Nelson and P.V. Preckel. The conditional beta distribution as a stochastic production function. *American Journal of Agricultural Economics*, 71(2):370–378, 1989.

[72] D. B. Nelson. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica: Journal of the Econometric Society*, 59(2):347–370, 1991.

[73] B. Norwood, M. C. Roberts, and J. L. Lusk. Ranking crop yield models using out-of-sample likelihood functions. *American Journal of Agricultural Economics*, 86, 2004.

[74] S. M. Nusser and J. J. Goebel. The National Resources Inventory: a long-term multi-resource monitoring programme. *Environmental and Ecological Statistics*, 4(3):181–204, 1997.

[75] OECD. Improving the environmental performance of agriculture: Policy options and market approaches. *Organisation for Economic Co-operation and Development*, 2001c.

[76] Nicholas Paulson and Bruce A. Babcock. Get a grip: Should area revenue coverage be offered through the farm bill or as a crop insurance program? *Working Paper 07-WP-440, CARD, ISU*, 2007.

[77] R. R. Picard and R. D. Cook. Cross-validation of regression models. *Journal of the American Statistical Association*, 79(387):575–583, 1984.

[78] N. E. Piggott and T. L. Marsh. Does food safety information impact US meat demand? *American Journal of Agricultural Economics*, 86(1):154, 2004.

[79] B. M. Potscher. Effects of model selection on inference. *Econometric Theory*, 7(2), 1991.

[80] S. Rabotyagov, T. Campbell, M. Jha, P. W. Gassman, J. Arnold, L. A. Kurkalova, S. Secchi, H. Feng, and C. L. Kling. Least-cost control of agricultural nutrient contributions to the Gulf of Mexico hypoxic zone. *Ecological Applications*, 20:1542–1555, 2010.

[81] M. C. Roberts, R. W. Mullen, and S. Prochaska. The Role of Yield-Price Correlation in Setting Optimal N Application Rates for Corn Production. In *2006 Annual meeting, July 23-26, Long Beach, CA*. American Agricultural Economics Association (New Name 2008: Agricultural and Applied Economics Association), 2006.

[82] V. Salin and N. H. Hooker. Stock market reaction to food recalls. *Review of Agricultural Economics*, 23(1):33, 2001.

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

[84] S. Secchi and B. A. Babcock. *Impact of high crop prices on environmental quality: A case of Iowa and the Conservation Reserve Program.* Center for Agricultural and Rural Development, Iowa State University, 2007.

[85] S. Secchi, P. W. Gassman, J. R. Williams, and B. A. Babcock. Corn-based ethanol production and environmental quality: A case of Iowa and the Conservation Reserve Program. *Environmental management*, 44(4):732–744, 2009.

[86] Silvia Secchi and Bruce A. Babcock. Impact of high corn prices on environmental quality: A case of iowa and the conservation reserve program. Technical Report 07-WP 447, Centre for Agricultural and Rural Development, Iowa State University, 2007.

[87] K. Shin and R. Pasupathy. An Algorithm for Fast Generation of Bivariate Poisson Random Vectors. *INFORMS Journal on Computing*, 22(1):81–92, 2010.

[88] Ricardo Smith-Ramirez. *On the Evaluation of Conservation Cost-Sharing Programs. An Appli-cation of a Monte Carlo EM Algorithm.* PhD thesis, University of Maryland, 2005.

[89] D. A. Sumner. Reducing Cotton Subsidies: The DDA Cotton Initiative. *Agricultural Trade Reform and the Doha Development Agenda*, 2006.

[90] K. Tanaka and J.J. Wu. Evaluating the effect of conservation policies on agricultural land use: a site-specific modeling approach. *Canadian Journal of Agricultural Economics*, 52(3):217–235, 2004.

[91] R. H. Thaler. Mental accounting matters. *Journal of Behavioral Decision Making*, 12(3):183–206, 1999.

[92] M. R. Thomsen and A. M. McKenzie. Market incentives for safe foods: an examination of shareholder losses from meat and poultry recalls. *American Journal of Agricultural Economics*, 83(3):526–537, 2001.

[93] P. K. Trivedi and D.M. Zimmer. *Copula modeling: an introduction for practitioners.* Now Publishers, 2007.

[94] USDA. Treatment of respiratory disease in u.s. feedlots. Veterinary services information sheet, USDA, Animal and Plant Health Inspection Service, Washington DC, October 2001.

[95] H. Van Ophem. A general method to estimate correlated discrete random variables. *Econometric Theory*, 15(2):228–237, 1999.

[96] M. R. Veall. Bootstrapping the process of model selection: An econometric example. *Journal of Applied Econometrics*, 7(1), 1992.

[97] M. R. Veall and K. F. Zimmermann. Pseudo-r^2 measures for some common limited dependent variable models. *Journal of Economic Surveys*, 10:241–59, 1996.

[98] Z. Wang, V. Salin, N. H. Hooker, and D. Leatham. Stock market reaction to food recalls: A GARCH application. *Applied Economics Letters*, 9(15):979–987, 2002.

[99] J. J. Wu. Crop Insurance, Acreage Decisions, and Nonpoint-Source Pollution. *American Journal of Agricultural Economics*, 81(2):305–320, 1999.

[100] J. J. Wu. Slippage effects of the conservation reserve program. *American Journal of Agricultural Economics*, 82(4):979–992, 2000.

[101] J. J. Wu and B. A. Babcock. Metamodeling Potential Nitrate Water Pollution in the Central United States. *Journal of Environmental Quality*, 28:1916–28, 1999.

[102] J. J. Wu and K. Segerson. The Impact of Policies and Land Characteristics on Potential Groundwater Pollution in Wisconsin. *American Journal of Agricultural Economics*, 77(4):1033–1047, 1995.

[103] J.J. Wu, R.M. Adams, C.L. Kling, and K. Tanaka. From microlevel decisions to landscape changes: an assessment of agricultural conservation policies. *American Journal of Agricultural Economics*, 86(1):26–41, 2004.

[104] F.A. Yates. External correspondence: Decompositions of the mean probability score. *Organizational Behavior and Human Performance*, 30, 1982.

[105] P. Zhang. Inference after variable selection in linear regression model. *Biometrika*, 79, 1992.

[106] J. Zhao, C. L. Kling, and L. Kurkalova. Alternative Green Payment Policies under Heterogeneity When Multiple Benefits Matter. *Working Paper 03-WP-341, CARD, ISU*, 2003.

[107] Y. Zheng, H. W. Kinnucan, and H. Thompson. News and volatility of food prices. *Applied Economics*, 40(13):1629–1635, 2008.