

# **Essays on Agricultural and Resource Economics**

By

Heesun Jang

A dissertation submitted in partial fulfillment of  
the requirements for the degree of

Doctor of Philosophy

(Agricultural and Applied Economics)

at the

UNIVERSITY OF WISCONSIN-MADISON

2015

Date of final oral examination: 05/11/2015

The dissertation is approved by the following members of the Final Oral Committee:

Xiaodong Du, Assistant Professor, Agricultural and Applied Economics

Jean-Paul Chavas, Anderson-Bascom Professor, Agricultural and Applied Economics

Brent Hueth, Associate Professor, Agricultural and Applied Economics

Dominic Parker, Assistant Professor, Agricultural and Applied Economics

Alan T. Sorensen, Professor, Economics

## Contents

List of Tables .....	iii
List of Figures .....	iii
Acknowledgements .....	v
Abstract of the Dissertation .....	vii
Chapter 1	
Price- and Policy- Induced Innovations: The Case of U.S. Biofuel .....	1
1. Introduction .....	1
2. Literature Review .....	4
3. Ethanol Patent Count Data and Descriptive Evidence .....	6
4. Empirical Methods .....	8
4.1. Knowledge Stock .....	9
4.2. Determinants of Innovation .....	11
5. Analysis of the Results .....	12
6. Conclusion .....	16
Chapter 2	
Evolving Techniques in Production Function Identification Illustrated in the Case of the US Dairy .....	27
1. Introduction .....	27
2. Empirical Framework .....	29
3. Data .....	37
4. Empirical Results .....	41
4.1. Estimation of the Production Function .....	41

4.2. Regional Productivity Decomposition .....	43
4.3. Driving Factors of the Productivity Growth .....	46
5. Conclusion .....	47
6. Appendix.....	62
Chapter 3	
An Empirical Structural Model of Productivity and Decision on the Conservation Reserve Program .....	65
1. Introduction.....	65
2. Background .....	68
3. Literature Review.....	70
4. Model .....	72
5. Estimation .....	74
5.1. The Output Supply Function.....	74
5.2. The Acreage Equations.....	76
6. Data.....	77
7. Empirical Results .....	80
7.1. Binary CRP Participation Decision.....	80
7.2. Optimal Output Supply and Land Allocations.....	81
7.3. Payment Premiums of High-Productivity Farms .....	82
8. Policy Simulation .....	83
9. Conclusion .....	87
10. Appendix.....	107
Reference .....	109

## List of Tables

Table 1-1. Ethanol Patents Granted to US Assignees, January 1976-June 2012. ....	18
Table 1-2. Regression Results for Probability of Being Cited of US Patents. ....	19
Table 1-3. Descriptive Statistics, 1978-2010. ....	20
Table 1-4. Regression Results for Determinants of Innovation. ....	21
Table 2-1. Descriptive Statistics of Production Regions: Sample Mean and Standard Deviation (in the parentheses). ....	49
Table 2-2. Descriptive Statistics of Major Inputs and Output: Sample Mean and Standard Deviation (in the parentheses). ....	50
Table 2-3. Descriptive Statistics of Entering, Exiting, and Surviving Dairy Farms (Averaged over 1987-2007). ....	51
Table 2-4. Descriptive Statistics of Entry and Exit Rate by Production Region. ....	52
Table 2-5. The OLS Regression Results. ....	53
Table 2-6. GNR Estimation Results. ....	54
Table 2-7. Control Function Approach Estimation Results in All Regions. ....	55
Table 2-8. Dynamic Olley-Oakes Productivity Decomposition Results. ....	56
Table 2-9. Descriptive Statistics of Farm- and Regional- Variables by Production Region. .....	57
Table 2-10. Marginal Effect in Probit Estimation of Exit. ....	58
Table 2-11. Determinants of Farm-Level Productivity. ....	59
Table 3-1. Descriptive Statistics of No-CRP, Partial, and Whole Farms. ....	90
Table 3-2. Descriptive Statistics of Partial Farms. ....	91
Table 3-3. Probit Regression Results. ....	92
Table 3-4. GMM Estimation Results. ....	93
Table 3-5. Productivity Regression Results. ....	94
Table 3-6. Prices Used in Simulation. ....	95
Table 3-7. Changes in CRP Acreage (in million acres). ....	96
Table 3-8. Additional CRP Acreage and Costs by Increased Payment (in million acres; \$ millions). ....	97

## List of Figures

Figure 1-1. Annual Ethanol Patent Counts by Category, 1975-2011. ....	22
Figure 1-2. Annual US Ethanol Patents by Assignee Group, 1975-2011.....	23
Figure 1-3. Average Probability of Being Cited by Patent Grant Year, 1977-2010.....	24
Figure 1-4. The Constructed Knowledge Stocks, 1977-2010.....	25
Figure 1-5. The Time Series of Included Explanatory Variables, 1978-2010.....	26
Figure 2-1. Farm Production Regions in the United States. ....	60
Figure 2-2. Average County-Level Temperature Humidity Index, 1987-2007.....	61
Figure 3-1. CRP Acreage and Corn Prices Received by Farmers, 1986-2013. ....	98
Figure 3-2. Histogram of Relative Productivity Estimates. ....	99
Figure 3-3. Actual Changes in CRP Acreage between 2007 and 2012 (million acres)..	100
Figure 3-4. Actual Changes in CRP Costs between 2007 and 2012 (\$ millions).....	101
Figure 3-5. Predicted Changes in CRP Acreage between 2007 and 2012 (million acres). .....	102
Figure 3-6. Additional CRP Acreage by 10% Increase in 2012 CRP Payment (million acres).....	103
Figure 3-7. Additional CRP Costs by 10% Increase in 2012 CRP Payment (\$ millions). .....	104
Figure 3-8. Additional CRP Acreage by 20% Increase in 2012 CRP Payment (million acres).....	105
Figure 3-9. Additional CRP Costs by 20% Increase in 2012 CRP Payment (\$ millions). .....	106

## Acknowledgements

This dissertation is the result of four years that have been rewarding in many aspects, but also challenging at times – which I think postgraduate studies should be. There are a number of people who helped my dissertation possible, and for this they deserve my gratitude.

First of all, I am greatly indebted to my advisor, Professor Xiaodong Du, for his guidance, support, and availability that constituted an important part of my dissertation. Working with him has been an enlightening and unforgettable experience. His continuous encouragement enabled me to accomplish things I thought were impossible.

My dissertation committee also deserves of many thanks. I thank Professor Jean-Paul Chavas for his valuable suggestions and insightful comments. He has always been willing to discuss and offer advice at key stages of my dissertation, and for that, I am grateful. I also thank Professor Brent Hueth, for his guidance, support, and for providing me a great opportunity to build my career and become a better researcher. I am grateful to Professor Dominic Parker for his challenging questions and comments that provided the inspiration for my third chapter. I also thank Professor Alan Sorensen for his insightful questions and comments for my second chapter.

I appreciate the Department of Agricultural and Applied Economics (AAE) had me as a student and provided great opportunities. I deeply thank each of members in the AAE Department, including professors, graduate students, and staff, for creating intellectually encouraging and home-like environment. I also acknowledge the USDA-NASS Wisconsin Field Office that supported me by providing data for my second and third chapters.

A special thanks to Professor Hojeong Park, my advisor at Korea University, whose support and advice were pivotal in my decision to pursue a Ph.D. I always remember and miss the moments I worked with him.

Last but not least, my deepest thanks to my family in Korea, for their love and unconditional support through all these times. Thanks so much for being at all times there for me no matter how far they are. I wouldn't be here today if there wasn't their love and constant dedication. Finally, I would like to express my gratitude to Youngdae, for his love and for being perfect companion through all these years. Without him, none of these would have been possible.

## **Abstract of the Dissertation**

My dissertation provides theoretical and empirical contributions to investigate the roles and sources of the technological innovation and the productivity growth via three essays.

Employing the patent count and citation data over 1977-2011, the first chapter explores the determining factors of innovative activities of the U.S. biofuel. We take into account both demand-side factors, such as crude oil price, government R&D expenditure on biofuel, and federal-level support policies, and supply-side factors represented by the constructed knowledge stocks, to quantify the effects on biofuel related innovations. The citation generation process is quantified using patent citation records and the estimates are used to construct the simple and weighted stocks of knowledge with weights of patent productivity. We confirm that both the demand and supply factors have positive and statistically significant effects on the technological innovations of biofuel in the United States.

The second chapter estimates the production technologies of major dairy producing regions in the US. The simultaneity and sample selection biases are carefully corrected following the recent development in the literature, especially Gandhi et al. (2013). The commonly applied control function approaches are found to be misleading in correcting for biases of coefficient estimates on flexible and quasi-fixed input variables. We show that the emerging dairy regions have relatively higher aggregate productivity than the traditional regions. Dynamic decomposition results indicate that surviving farms play a more important role in the regional productivity growth than entering and exiting farms. Farm and regional driving forces of farm productivity are also examined.

The third chapter investigates landowners' decisions on the Conservation Reserve Program (CRP) under a model of land uses. CRP provides financial incentives for landowners to idle erodible and marginal farmland, which has contributed to a number of environmental benefits that otherwise would have not been achieved. For the last few years, however, CRP has faced a number of changes, primarily as record-high crop prices significantly affected landowners' interests in the program. Although there is an extensive literature on CRP, we lack evidence on how landowners react to changes in agricultural market conditions and CRP rental rates and hence how it affects the program enrollment and costs. In this chapter, we attempt to investigate landowners' incentives to participate in CRP, focusing on the linkage between landowner's bid on CRP payment and the unobserved agricultural productivity. We develop and estimate an empirical structural model to examine the manner in which agricultural productivity, market conditions, and CRP payment affect landowners' land use decisions. A novel identification strategy is employed to control for endogeneity of CRP payment and landowners' self-selection into the program. The parameter estimates are used to simulate how increased agricultural prices and CRP payment influence the program enrollment and costs.

# Chapter 1

## Price- and Policy- Induced Innovations: The Case of U.S. Biofuel<sup>1</sup>

### 1. Introduction

Finite and unequal distribution of fossil fuel resources, together with growing concerns on environment, has brought significant changes to world energy production and consumption. Countries worldwide facing the issues of energy security and low carbon economy have been stimulated to seek alternative energy resources to displace fossil fuel sources. As a successful example, ethanol is the most widely used biofuel in the world. It has been grabbing significant attention in some countries including U.S. and Brazil. These two countries accounted for about 88% of worldwide ethanol production in 2010.

The US ethanol production experienced rapid growth in the late 1970s after the subsidy established by the Energy Policy Act of 1978 launched the industry. The domestic ethanol producers were also protected from Brazilian competitors by the import tariff since 1980s.<sup>2</sup> From 2005 to 2011, ethanol production has quadrupled from 3.9 billion gallons to 13.9 billion gallons currently substituting about 10% of the U.S.

---

<sup>1</sup> This chapter was published in 2013 in the *Journal of Agricultural and Resource Economics* 38(3): 299-311, titled as “Price- and Policy-Induced Innovations: The Case of U.S. Biofuel.”

<sup>2</sup> The original federal support policies were revised by later legislation, but the basic structure of tax and import tariff had remained in place until the end of 2011 (see California Energy Commission 2004 for a more detailed discussion).

gasoline supply (RFA 2012). Substantial expansion of ethanol production and resulting feedstock demand raise some concerns including, for example, increasing crop price, intensive corn production and potentially significant land use changes as well as associated environmental impacts.

The growth of U.S. ethanol industry is largely accelerated by the federal and state government support policies. The major federal-level support policies in recent years include (i) the blending mandate, which was established by the Renewable Fuel Standard (RFS) in 2005 and further expanded by the Energy Independence and Security Act of 2007, that requires blending of minimum amount of ethanol into fuel gasoline each year, (ii) the tax credit of \$0.45/gallon for ethanol blenders, and (iii) the import tariff of \$0.54/gallon. As the tax credit and import tariff were expired at the end of 2011, the U.S. ethanol industry has grown into an established and mature industry. Production costs of corn ethanol dropped by about 65% and industrial processing costs decreased by approximately 45% between 1983 and 2005 (Hettinga et al. 2009; Chen and Khanna 2012). Over the same period, we see upward trend in patent granting of ethanol-related innovations. There were massive applications of patents between late 1970s and early 1980s. After a quiet development period, a large number of technological innovations with applications of new patents emerged after 2005.

In general, technological innovations are pushed forward by both demand and supply factors (see Verholini and Galeotti 2011 for a more detailed discussion). The supply factors refer to changes affecting existing knowledge or potential opportunities for future inventive activities. The demand-side factors make new innovations more or less profitable for a given level of existing knowledge. This study attempts to provide a

rigorous empirical understanding on the determinants of technological innovations of the U.S. ethanol industry. Several unique features of the U.S. biofuel innovations motivate this study. First, the booms of biofuel innovations in the late 1970's and after 2005 are largely stimulated by not only demand-side but also supply-side factors including oil price, R&D expenditure, government policies, and existing knowledge. Similar to the findings in the literature, e.g., Popp (2002), we find strong and significant impact of oil price and existing knowledge on biofuel innovations.

Second, the U.S. ethanol industry provides an interesting setting within which to examine the driving forces of innovative activities and to test the hypotheses of induced innovations first proposed in Popp (2002). Insights into the ethanol industry will provide lessons on various aspects, e.g., capital investment and policy intervention, of the development of other renewable energy resources. In addition, technology evolution of ethanol should present valuable information on developing mechanisms to stimulate future renewable energy innovations.

Finally, utilization of patent data by itself should be interesting. Patent counts and citations provide not only the quantity but also the scientific value of innovative output. Patent citations indicate knowledge contribution by prior research, or in other words, the usefulness of existing knowledge to later inventors. Following the method developed in Popp (2002), we construct the stocks of knowledge measuring the level of current technological knowledge on ethanol, which is considered as the supply-side determining factor of biofuel innovations. It turns out that incorporating the measure of existing knowledge is important for quantifying the price and policy effects.

Our contribution is three-fold. First, we confirm that not only the demand for knowledge innovation and accumulation, but also the supply-side factor represented by the existing knowledge stock plays an important role in the innovation process. Second, both simple and weighted knowledge stocks reflect the existing knowledge of U.S. ethanol and have positive and significant impact on ethanol innovations. Finally, empirical analysis of the innovation determinants finds that both demand and supply factors positively and significantly affect ethanol innovations.

The rest of paper is organized as follows. Section 2 reviews the existing literature. Patent datasets used in the study are discussed in Section 3. In Section 4 we describe the empirical models for constructing the knowledge stocks and for quantifying the determining factors of technological innovation. Section 5 analyzes the estimation results followed by some concluding comments in the final section.

## **2. Literature Review**

A number of studies employ patent data to analyze driving forces of technological innovations in energy and renewable energy sectors. Popp (2002) is the first study proposing the concept of knowledge stock in studying energy-related innovations using patent data. The study shows that (i) both energy prices and existing knowledge stock have positive and significant effects on energy-efficient innovations, and (ii) ignoring quality of the knowledge biases the estimates. Accounting for both domestic and international knowledge flows, Verdolini and Galeotti (2011) construct internal and external knowledge stocks to study determinants of innovation in energy technologies. Using the data of 38 innovating countries from 1975 to 2000, the authors find that higher

geographical and technological distances are associated with lower probability of knowledge flows.

Johnstone, Hascic and Popp (2008) examine the effects of policy incentives on the patent counts of a number of renewable energy technologies. Their sample includes 25 high-income countries that have adopted various support policies to encourage the development of renewable energy resources. While public policy plays a significant role in inducing innovations, the efficacy of alternative policy instruments varies by energy sources. De Freitas and Kaneko (2012) confirm a unidirectional relation from consumption to technological innovation of sugarcane ethanol in Brazil.

The studies on the innovative activities and the determinants of U.S. ethanol innovations are sparse. Hettinga et al. (2009) investigate technological development of U.S. ethanol production using an experience curve approach where technological learning is identified in two separate systems: corn production and ethanol processing. Main drivers behind large cost reductions include high ethanol yields and the introduction of ethanol-specific automated technologies. Using data on U.S. dry-mill ethanol processing costs of 1983-2005, Chen and Khanna (2012) investigate the reasons underlying the declining processing costs. They find that (i) U.S. corn ethanol production exhibits decreasing return-to-scale, (ii) learning-by-doing plays an important role in cost reduction, and (iii) imports of sugarcane ethanol contribute to increasing competitiveness of the domestic industry. The most relevant to our work is Karmarkar-Deshmukh and Pray (2009). Their results indicate significant positive effects of oil prices and federal research grants on ethanol-related innovations. The study was conducted in 2008 and 445 ethanol-

related patents are identified and utilized. They didn't account for the effect of knowledge stock by assuming patents are equally important for later innovations.

The U.S. government has been playing an important role in stimulating biofuel development in recent years by providing significant amount of R&D funding and strong policy incentives. However, the effect of government policies on technological innovations is not found to be clear in the literature. For example, in studying the innovations related to biomass for electricity, Johnstone, Hascic and Popp (2008) find that government R&D spending has a negative and significant effect on patent applications, implying that government R&D funding crowds out private spending. For U.S. corn ethanol, Karmarkar-Deshmukh and Pray (2009) show that the research funding provided by the federal government has a positive and significant effect on technological innovations. But they also show that the ethanol tax credit has a negative and significant effect on ethanol innovations, while the effect of mandate policy was negative but insignificant.

### **3. Ethanol Patent Count Data and Descriptive Evidence**

In this study, technological innovations of ethanol are represented by the number of patents registered at the U.S. Patent and Trademark Office (USPTO). To identify the ethanol-related patents, we search the USPTO database and collect the patents having the word "ethanol" in title or abstract. The initial search generated a total of 3,539 patents applied over 1975-2011 and granted from January 1976 to June 2012. Patent descriptions are then manually reviewed and screened for direct relevance. The screening yields 1,090 ethanol patents over the sample period.

Following de Freitas and Kaneko (2012), we classify the ethanol patents into five categories including agricultural feedstock production, ethanol production process, engine and ethanol combustion, ethanol by-product, and emission treatment.<sup>3</sup> Figure 1 presents the annual patent counts by category sorted by application year. There are a total of 1,090 ethanol patents registered at USPTO, 673 of which are granted to U.S. assignees and are the focus of the current study.<sup>4</sup>

As shown in Figure 1, the trend of patenting activity is largely consistent with the development of U.S. ethanol industry. Patent applications reached the peak in early 1980s and declined to around 20 each year after that. Applications grew by more than 20% per year since 2005, which was largely triggered by increasing domestic ethanol demand and high crude oil prices. Different from the Brazil's ethanol industry where agriculture-related patents experienced a major increase during the 2000s (de Freitas and Kaneko 2012), patenting activities in the U.S. are dominated by the innovations related to ethanol production and engine/combustion.

Figure 2 illustrates the trend of ethanol patents by U.S. assignee groups from 1975 to 2011.<sup>5</sup> There are three assignee groups including the government, universities, and

---

<sup>3</sup> We exclude the category of transport and storage included in de Freitas and Kaneko (2012) as there are no relevant patents in the U.S. data.

<sup>4</sup> Notice that an assignee and an inventor of a patent may be different in some cases because that, for example, it is more likely for an employee to assign a patent to a company. For the patents with U.S. assignees, which are the focus of our study, approximately over 90% of inventors are resident in the US. So we don't make the distinction between assignee and inventor here as it has little effect on patent counts and the latter empirical results.

<sup>5</sup> Patents assigned to more than one assignee are counted multiple times. This yields a total of 1,090 patents by assignees.

private companies. Over the sample period, private companies led the ethanol-related technological innovations followed by universities and the government. Figure 2 shows that the U.S. government was more active in patenting than universities during early 1980s. However, universities contribute more to the recent ethanol boom after 2005. We consider the total US patents in the latter analysis because that government support policies play a critical role in the development of U.S. ethanol industry and related technological innovations. Not only private companies but also universities and government-related research organizations are significantly affected by the extensive support and make important contribution to ethanol related knowledge accumulation. Including patents of all groups paints a more complete picture of the knowledge creation and accumulation processes.

Table 1 summarizes the information on patent citations, the distribution of which is highly skewed. About 38% of U.S. patents received no citation over the whole sample period. 37% of the patents received one to ten citations and 16% received eleven to thirty citations. Only the remaining 3% received thirty or more citations. On average, one patent is cited approximately seven times over the past 37 years.<sup>6</sup>

#### **4. Empirical Methods**

In this section we describe the empirical methods for (i) constructing knowledge stocks, and (ii) analyzing driving factors of technological innovations of the U.S. biofuel.

---

<sup>6</sup> Note that a significant number of patents are applied and granted toward the end of the sample period and didn't get time to generate as many citations as earlier patents. This may have some implications for the estimates of the citation generation process explained in a later section.

## 4.1. Knowledge Stock

Knowledge stock is a commonly applied measure of existing stock of scientific knowledge (see, e.g., Jaffe and Trajtenberg 1996; Popp 2002). Higher level of existing knowledge stock indicates better technological opportunities and should push forward more innovative activities in future periods. Patent citations contain useful information for knowledge stock construction. Citation to a previously granted patent reflects the usefulness of the earlier cited patent and its knowledge contribution to the new citing patent (Popp 2002). Hence, frequent citations suggest high quality of the cited patent and its more important contribution to knowledge accumulation of the industry.

Following the literature, we use patent citations to estimate the parameters underlying the citation generation process after taking into account relative importance of individual patents. The estimated parameters are then combined with patent counts to construct the time-varying knowledge stocks of ethanol-related technologies.

The probability that a patent applied in year  $T$  cites a patent  $k$  granted in year  $t$  can be described by the following double-exponential function:<sup>7</sup>

$$(1) \quad p(k, K) = \alpha(k, K) \exp[-\beta_1(T-t)][1 - \exp(-\beta_2(T-t))]$$

where knowledge becomes obsolete at the rate of  $\beta_1$  and diffuses by the rate of  $\beta_2$ .

$\alpha(k, K)$  denotes the parameters capturing the attributes of the citing or cited patents that may influence the probability of citation. The citation probability,  $p(k, K)$ , which is also referred as citation frequency, is defined as a function of the citation lag  $(T-t)$  and the related parameters.

---

<sup>7</sup> Eqn. (5), p. 168, Popp (2002).

Denote the number of citations in a cited-citing year pair as  $c_{k,K}$ , cited patents granted in year  $t$  as  $n_t$ , and citing patents applied in year  $T$  as  $n_T$ . Eqn. (1) can be rewritten as:<sup>8,9</sup>

$$(2) \quad p_{k,K} \equiv \frac{c_{k,K}}{n_t n_T} = \alpha_0 \alpha_t \alpha_T \exp[-\beta_1(T-t)][1 - \exp(-\beta_2(T-t))]$$

where  $\alpha(k, K)$  in Eqn. (1) incorporates the constant  $\alpha_0$ , cited year effect  $\alpha_t$  and citing year effect  $\alpha_T$ . The cited year effect  $\alpha_t$  indicates the usefulness of the cited patents. Higher cited year effect indicates higher probability of being cited and thus reflects the relative importance of the cited patent.

Two types of knowledge stocks, simple ( $K_t^{simple}$ ) and weighted knowledge stock ( $K_t^{weighted}$ ), for year  $t$  are then constructed using the estimated parameters in Eqn. (2) and the patent counts as follows:

$$(3) \quad K_t^{simple} = \sum_{s=1977}^t Pat_s \exp[-\hat{\beta}_1(t-s)][1 - \exp(-\hat{\beta}_2(t-s))]$$

$$(4) \quad K_t^{weighted} = \sum_{s=1977}^t \hat{\alpha}_s Pat_s \exp[-\hat{\beta}_1(t-s)][1 - \exp(-\hat{\beta}_2(t-s))]$$

where  $Pat_s$  represents the number of patents granted in year  $s$ ,  $\hat{\alpha}_s$  is the estimated cited year effect, and  $\hat{\beta}_1$  and  $\hat{\beta}_2$  represent the estimated decay and diffusion rates, respectively.

The simple knowledge stock ( $K_t^{simple}$ ) provides a measure of existing knowledge

---

<sup>8</sup> Eqn. (6), p. 169, Popp (2002).

<sup>9</sup> We excluded the observations at the year when  $t = T$  as such citation pairs are rare in the data.

accumulation of ethanol using annual patent counts. The weighted knowledge stock ( $K_t^{weighted}$ ) quantifies the knowledge stock after incorporating patents' relative usefulness by using the cited year effect as a multiplicative factor. In our study, both knowledge stocks are constructed over the period of 1977-2010.

## 4.2. Determinants of Innovation

We proceed to analyze the determinants of the U.S. biofuel innovations. The constructed knowledge stocks are included to represent the supply-side factor. Crude oil prices, government R&D expenditure on biofuel and support policy indicators are considered to be the demand-side factors.

In this study, the ethanol-related US patents applied in year  $t$ ,  $Pat_t$ , are determined as:<sup>10</sup>

$$(5) \quad Pat_t = \exp\left[\gamma_1 K_{t-1} + \gamma_2 P_t^E + \gamma_3 R \& D_t + \gamma_4 Policy_t + \varepsilon_t\right]$$

where  $K_{t-1}$  is the lagged simple (or weighted) knowledge stock of year  $t-1$ .  $P_t^E$  is the annual imported crude oil price deflated by consumer price index.  $R \& D_t$  denotes the real government R&D expenditure on biofuel (in 2012 dollars).<sup>11</sup>  $Policy_t$  is the policy

---

<sup>10</sup> As known in the literature (see, e.g., Aghion et al. 2005), there exists a strong relation between product market competition and innovation. In this study, we assume a perfectly competitive biofuel market and don't account for the possible impact of market structure on innovations.

<sup>11</sup> The data is part of the government energy technology R&D budgets collected by the International Energy Agency (IEA) and can be downloaded from the IEA website <http://www.iea.org/statistics/RDDonlinedataservice/>. We thank an anonymous reviewer for pointing this out. The database also includes R&D budgets for other renewable energy resources including solar and wind, which are used as instrumental variables in a later section.

indicator for the major ethanol support policies of the U.S. federal government after 2005, and  $\varepsilon_t$  stands for unobservable errors with zero mean. We use two policy proxy variables: (i) a dummy variable that takes the value of one for 2006 and after indicating the presence of major ethanol support policies (named “Policy dummy”), and (ii) annual volume mandate of ethanol under the RFS (and revised RFS2) since 2006 (named “RFS mandate”). We assume that the policies take effect one year after establishment. While the “RFS mandate” attempts to pick up mainly the effect of the blending mandate, the “Policy dummy” is to capture the aggregate effect of all existing federal-level support policies.

## 5. Analysis of the Results

First, we discuss the results on the citation generation process specified in Eqn. (2) and the knowledge stock construction in Eqns. (3) and (4). We use ethanol patents granted over 1977-2010 for the cited group ( $k$ ). The citing patents ( $K$ ) are the ones that cite patents in the group of  $k$  and are sorted by the application years of 1978-2011. A total of 6,989 pairs are included after sorted by the cited-citing years. Eqn. (2) is then estimated using nonlinear least squares method with White’s heteroskedasticity-consistent standard errors (White 1980).<sup>12</sup> It is difficult for the estimates to converge with separated parameters  $\alpha_t$  and  $\alpha_T$  for each cited-citing year, therefore we group the cited-citing

---

<sup>12</sup> Note that in Eqn. (2) each observation is the average probability of citation of an individual cohort, i.e., average of a group of patents. The averages tend to be more accurate for larger groups. So there exists potential heteroskedasticity problem that is the reason why we apply the heteroskedasticity-consistent estimator. Thank an anonymous reviewer for pointing this out.

years into three- by three-year intervals assuming that  $\alpha_t$  and  $\alpha_T$  are constant over the three years within the interval, but can vary across the intervals.

Table 2 reports the estimation results of the citation generation process. If we normalize the estimated cited year effect in the initial year of 1977,  $\hat{\alpha}_{1977}$  ( $=3.97$ ), to one, the estimated values of a later year,  $\hat{\alpha}_t$ , can be interpreted as relative to the base year. For example, the estimate of  $\hat{\alpha}_{1978-1980}$  becomes 1.78 ( $=7.08/3.97$ ), which means that the likelihood for a patent issued in 1978-1980 to receive a citation is about 78% higher than that of a patent issued in 1977. In Table 2, the cited year effects  $\hat{\alpha}_t$  decrease over 1978-2010, while the citing year effects  $\hat{\alpha}_T$  increase over 1982-2011. This is reasonable given the fact that the earlier a patent is granted, the higher chance it will be cited. Likewise, after the earlier ethanol boom, the later a patent is applied, the higher chance it cites other earlier granted patents. This is also consistent with the probability of citation ( $p_{k,K} \equiv c_{k,K} / n_t n_T$ ), which is calculated from data and presented in Figure 3. The citation probability was high before 1980, stable between 1980 and 2005, and increase significantly after 2005. The relatively higher probability of citation before 1980 is captured by the higher cited year effects, and that after 2005 is reflected into higher citing year effects as a significant number of citations are generated by newly applied patents during the recent ethanol boom. Both the estimated decay rate of  $\beta_1$  ( $=0.42$ ) and the diffusion rate of  $\beta_2$  ( $=0.0002$ ) are of comparable magnitude to those of Popp (2002) where the two estimates are 0.353 and 0.00199, respectively.

Figure 4 plots the constructed simple and weighted knowledge stocks over the period of 1977-2010. Both knowledge stocks reach the peak in mid-1980s. While

weighted knowledge stock declines after that, simple knowledge stock shows an upward trend since 2005, which indicates a significant increase of recent ethanol-related innovations. The falling value of weighted knowledge stock suggests deteriorating quality of knowledge represented by patents over time, or a decreasing rate of return of research (Popp 2002).

Time series of the variables included in the estimation of innovation determinants are presented in Figure 5 and Table 3 reports the corresponding descriptive statistics. Except the concave shape of the weighted knowledge stock, real crude oil prices and government biofuel R&D expenditure show consistent patterns with patenting activities over time. Peaks of patent application appear around 1980 and after 2005, together with relatively high crude oil prices and government R&D expenditure. Eqn. (5) is estimated using a Negative Binomial (NB) model as the Lagrange Multiplier (LM) test of over-dispersion rejects the Poisson model.

Government R&D expenditure on biofuel is considered to be endogenous in this setting as it could be potentially determined by some other factors influencing patenting activities at the same time. To deal with the endogeneity issue we rely on valid instrument variables (IVs), which are required to be directly related to biofuel R&D expenditure but affect patent generation only through R&D. We use government R&D expenditures on wind and solar as the instrumental variables because it is reasonable to assume that (i) government funding decisions for various renewable sources, such as wind, solar and biofuel, are closely correlated, and (ii) government R&D expenditures on

wind or solar have not effect on innovative activities of biofuel.<sup>13</sup> The first stage regression results are summarized in the lower panel of Table 4, which confirm the validity of the chosen IVs.

To address the endogeneity concern of knowledge stocks, following Popp (2002) we use a time trend and lagged GDP as instrumental variables for both simple and weighted knowledge stocks. In the NB regression, we explore a two-stage procedure. In the first stage, the endogenous biofuel R&D expenditure and knowledge stocks are regressed on the chosen IVs and other exogenous variables including oil price. The predicted values of the first stage regression are then used as regressors in the second stage NB regression. The standard errors are obtained using the standard bootstrapping method.<sup>14</sup>

The upper panel of Table 4 presents the estimation results of Eqn. (5). Columns I and II (III and IV) report the estimation results for including the lagged weighted (or simple) knowledge stock, which is omitted in Columns V and VI. As expected the results indicate positive and statistically significant effects of one-year lagged weighted (or simple) knowledge stock and crude oil prices on the technological innovations of the U.S. ethanol industry. Furthermore, it is found to be necessary to incorporate the knowledge

---

<sup>13</sup> We thank an anonymous reviewer for suggesting the IVs. As discussed above, the government expenditures for wind and solar are also obtained from the IEA website.

<sup>14</sup> Notice that the first stage regression results for the knowledge stocks presented in the lower panel of Table 4 indicate that the IVs are weak given the relatively low values of F statistics, especially the simple knowledge stock. As a robustness check, we run the NB regressions with the knowledge stock not being instrumented. The results are largely consistent with that reported in Columns I-IV of Table 4 except that the estimated coefficients on the weighted knowledge stock are positive but not statistically significant at 10% level.

stock, the supply-side factor. The biofuel R&D expenditures and policy variables are not found to be important technology-push factors except that the coefficient of biofuel R&D in Column II is statistically significant at 10% level. Expressing the estimated coefficients as incidence rate ratios (for the specification in Column I),<sup>15</sup> we find that one unit increase in the weighted knowledge stock is associated with about 2.2% increase of number of ethanol patents; one unit increase in the crude oil price increases the probability of innovation by about 4.0%; and one unit increase in the government R&D expenditure increases the probability of innovation by 2.3%.

We notice the difference between the two policy variables that the RFS mandate variable attempts to capture the effect of blending mandate only, the annual dummies after 2006 account for the effect of all ethanol support policies implemented through the period. There may be other reasons why the policy effect is not found to be significant. For example, the support policies only exist by the end of the sample period, which are coincident with high crude oil price and government R&D expenditure, therefore the policy effect may be partially captured by other variables.

## 6. Conclusion

Employing the patent count and citation data of ethanol-related technologies, this study extends the analysis of induced innovation to the U.S. biofuel, which has grown to a

---

<sup>15</sup> The estimates are transformed from  $\hat{\beta}$  to  $\exp(\hat{\beta})$  together with standard errors.  $\hat{\beta}$  can be interpreted as the estimated rate ratio for a one unit increase in the explanatory variable  $X_{\hat{\beta}}$  holding the other variables constant. In other words, if the level of  $X_{\hat{\beta}}$  increases by one unit, the rate for patents would be expected to change by a factor of  $\hat{\beta}$  (increases if  $\hat{\beta} > 1$  or decreases if  $\hat{\beta} < 1$ ) (Stata Corporation 2001). This is implemented using the `-irr` option in the STATA command. The results are available upon request.

mature industry through intensive federal and state government support policies as well as R&D funding.

Following the literature, we incorporate the effect of knowledge accumulation by quantifying the citation generation process and constructing the existing knowledge stocks. The empirical analysis of the determinants of technological innovation finds that both supply and demand factors, including knowledge stocks and crude oil price, positively and significantly affect ethanol-related innovations.

Our research suggests the need for future research in the following areas. First, availability of more satisfactory variables to proxy the demand-side factors such as policies and private R&D expenditure on biofuel would further strengthen the results in the paper. Second, it would be interesting to examine how the geographic characteristics of patent counts and citations differ across assignee groups. Third, as Brazil is the second largest ethanol producing country, it would generate meaningful policy discussion if we can directly extend the study to the sugarcane ethanol innovations in Brazil.

Table 1-1. Ethanol Patents Granted to US Assignees, January 1976-June 2012.

Total US Patents	673
Cited patents	418 (62%)
Average citation per patent	7.07
<hr/>	
<i>Citations</i>	<i># (%)</i>
0	255(38%)
1-10	252(37%)
11-20	105(16%)
21-30	38(6%)
>30	23(3%)
Max	91

Table 1-2. Regression Results for Probability of Being Cited of US Patents.

Variable	Estimates	Variable	Estimates
Cited Year Effects $\alpha_i$		Citing Year Effects $\alpha_r$	
1977	3.97 <sup>b</sup> (1.81)	1978	5.01 (8.47)
1978-1980	7.08 <sup>c</sup> (1.09)	1979-1981	1.09 (1.62)
1981-1983	5.67 <sup>c</sup> (0.58)	1982-1984	0.44 (0.54)
1984-1986	4.19 <sup>c</sup> (1.04)	1985-1987	0.44 (0.54)
1987-1989	1.99 <sup>b</sup> (0.84)	1988-1990	0.57 (0.62)
1990-1992	1.41 <sup>a</sup> (0.85)	1991-1993	1.08 (0.96)
1993-1995	0.82 (0.64)	1994-1996	1.75 (1.23)
1996-1998	0.53 (0.51)	1997-1999	2.82 <sup>a</sup> (1.47)
1999-2001	0.34 (0.39)	2000-2002	3.45 <sup>c</sup> (1.24)
2002-2004	0.20 (0.26)	2003-2005	4.44 <sup>c</sup> (0.89)
2005-2007	0.13 (0.20)	2006-2008	8.17 <sup>c</sup> (0.89)
2008-2010	0.16 (0.27)	2009-2011	15.92 <sup>c</sup> (3.50)
Decay rate ( $\beta_1$ )	0.42 <sup>c</sup> (0.07)		
Diffusion rate ( $\beta_2$ )	0.0002 (0.0003)		
Constant ( $\alpha_0$ )	1.49 <sup>c</sup> (0.0001)		
R squared	0.53		

Note:

(1) Standard errors are in the parentheses.

(2) The superscripts, a, b, and c, denote significance at 0.10, 0.05, and 0.01 levels, respectively.

Table 1-3. Descriptive Statistics, 1978-2010.

Variable	Observations	Mean	Std. Dev.	Min	Max
Patents	33	20.67	15.74	3	58
Weighted knowledge stock	33	24.97	27.43	1	85.11
Simple knowledge stock	33	33.95	15.36	1	63.71
Deflated Oil Price (\$/barrel)	33	48.16	23.72	16.97	98.61
Biofuel R&D (million \$)	33	145.25	239.25	0.74	1235.02

Note: Simple and weighted knowledge stocks are normalized using 1978 as the base year, i.e., both stocks in 1978 are equal to 1.

Table 1-4. Regression Results for Determinants of Innovation.

Variable	I	II	III	IV	V	VI
Lagged	0.022 <sup>c</sup>	0.030 <sup>c</sup>				
Weighted Stock	(0.006)	(0.008)				
Lagged			0.040 <sup>c</sup>	0.041 <sup>c</sup>		
Simple Stock			(0.004)	(0.004)		
Oil Price	0.039 <sup>c</sup>	0.031 <sup>c</sup>	0.028 <sup>c</sup>	0.027 <sup>c</sup>	0.052 <sup>a</sup>	0.052 <sup>c</sup>
	(0.008)	(0.01)	(0.004)	(0.005)	(0.006)	(0.005)
Biofuel R&D	0.023	0.027 <sup>a</sup>	0.008	0.007	0.025 <sup>a</sup>	0.021
	(0.02)	(0.015)	(0.007)	(0.008)	(0.013)	(0.013)
RFS mandate	0.024		0.018		-0.102 <sup>a</sup>	
	(0.12)		(0.048)		(0.053)	
Policy dummy		0.977		0.33		-0.67
		(0.72)		(0.39)		(0.49)
Log likelihood	-123.64	-122.34	-101.63	-100.27	-135.19	-135.96
$\chi^2$	300.83	881.49	1888.70	1238.52	359.14	307.09
<i>First stage regression</i>	Biofuel R&D	Weighted Stock	Simple Stock			
Wind R&D	6.19 <sup>c</sup>					
	(1.12)					
Solar R&D	-1.02 <sup>c</sup>					
	(0.23)					
Oil Price	0.26	-0.11	-0.07			
	(0.16)	(0.19)	(0.13)			
Lagged GDP		-1.77	-1.25			
		(1.35)	(0.91)			
Time trend		2.85	3.47			
		(3.67)	(2.47)			
Constant	-8.60	-5492.23	-6777.01			
	(6.89)	(7201.69)	(4847.94)			
R-squared	0.56	0.48	0.15			
F statistics	12.27	8.75	1.60			

Note:

(1) Bootstrapping standard errors are in the parentheses.

(2) Simple and weighted knowledge stocks are normalized using 1978 as the base year.

(3) The superscripts, a, b, and c, denote significance at 0.10, 0.05, and 0.01 levels, respectively.

Figure 1-1. Annual Ethanol Patent Counts by Category, 1975-2011.

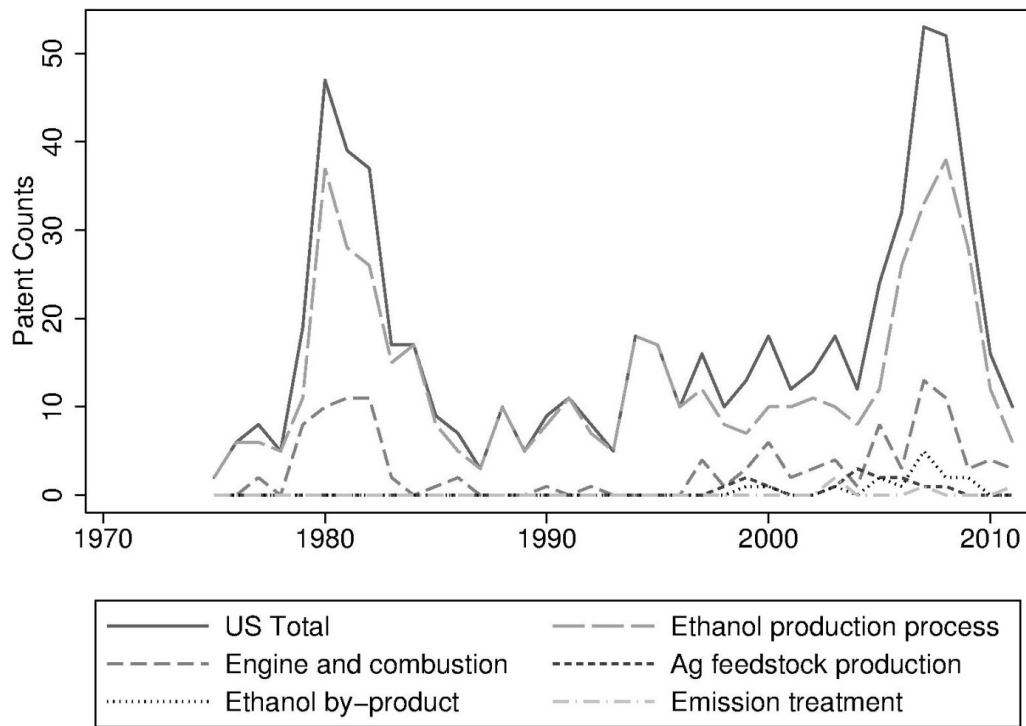


Figure 1-2. Annual US Ethanol Patents by Assignee Group, 1975-2011.

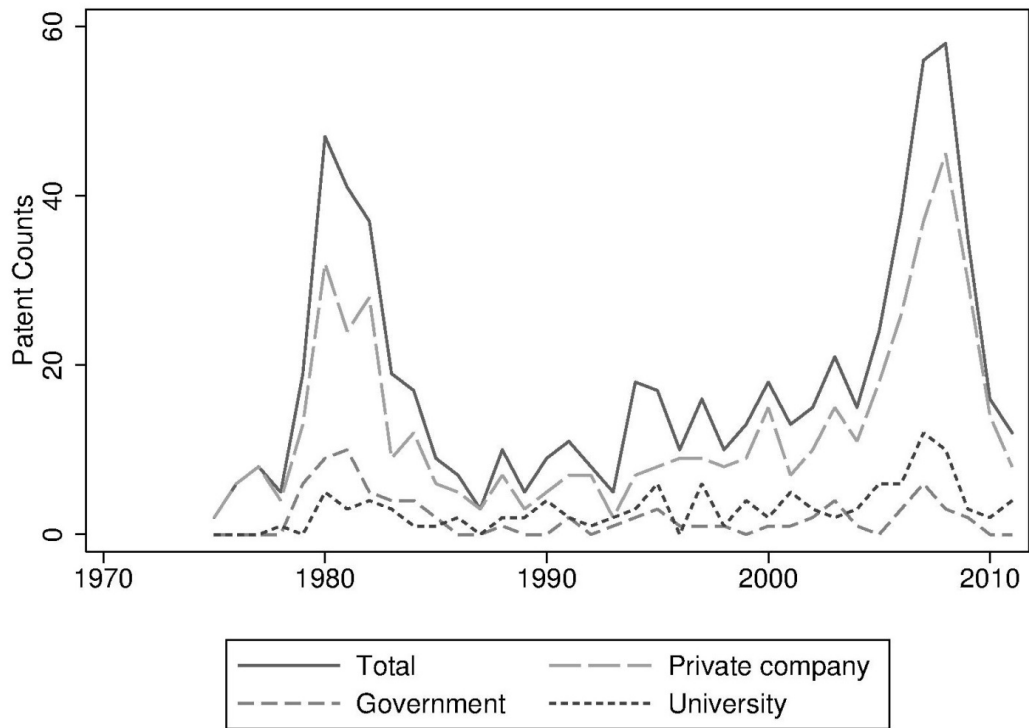


Figure 1-3. Average Probability of Being Cited by Patent Grant Year, 1977-2010.

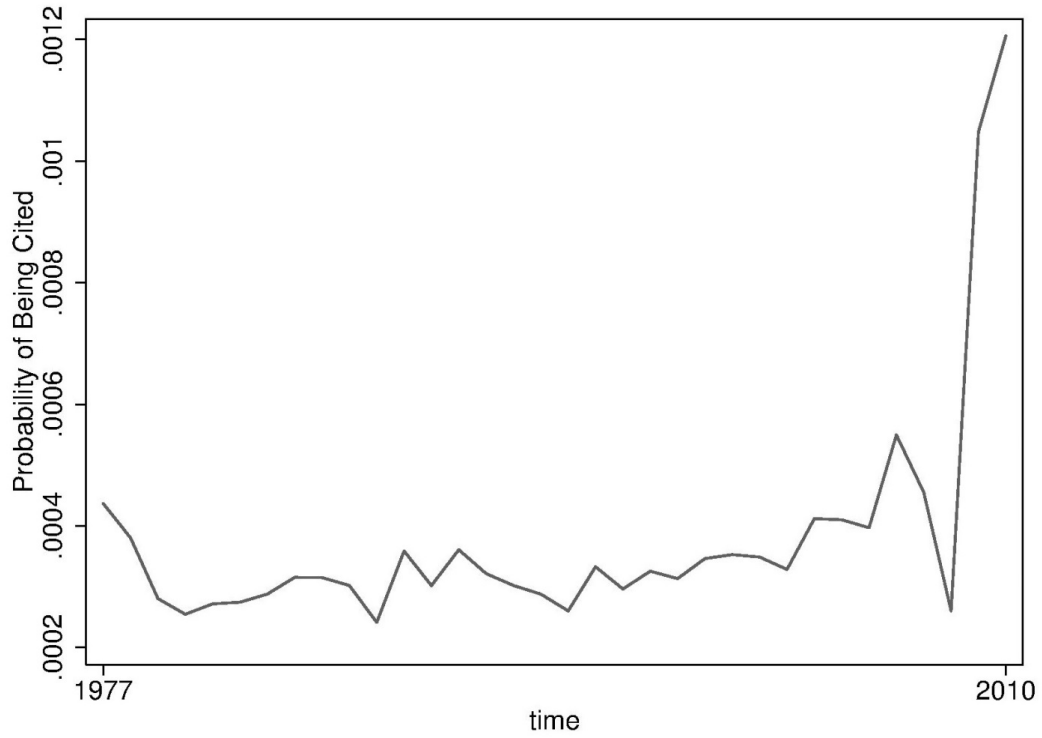


Figure 1-4. The Constructed Knowledge Stocks, 1977-2010.

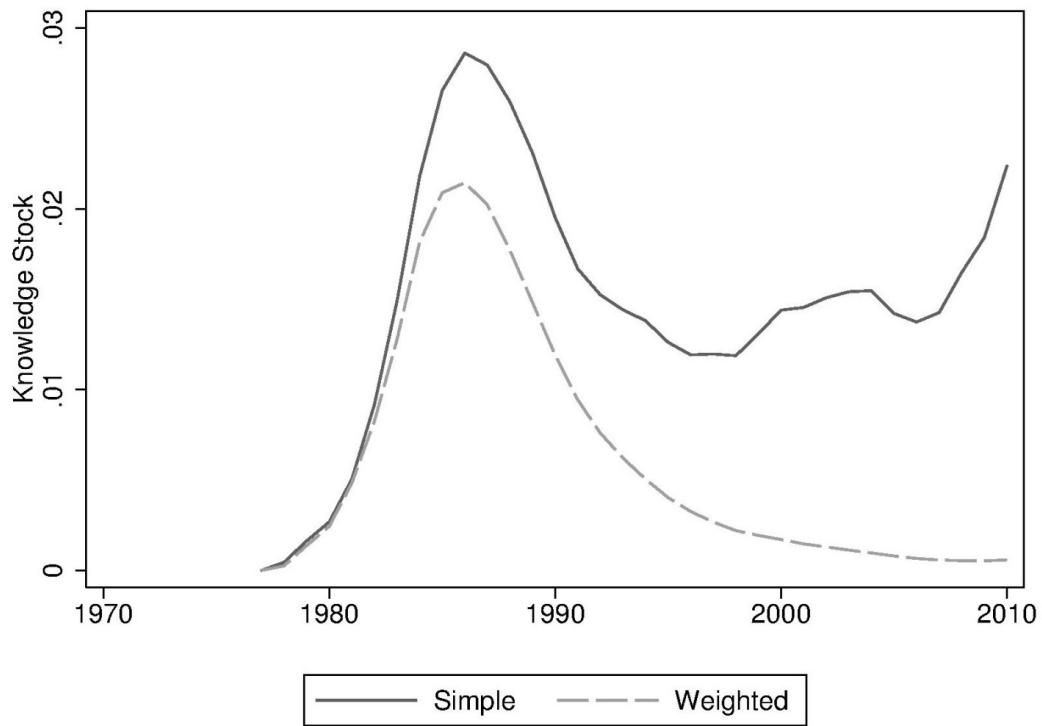
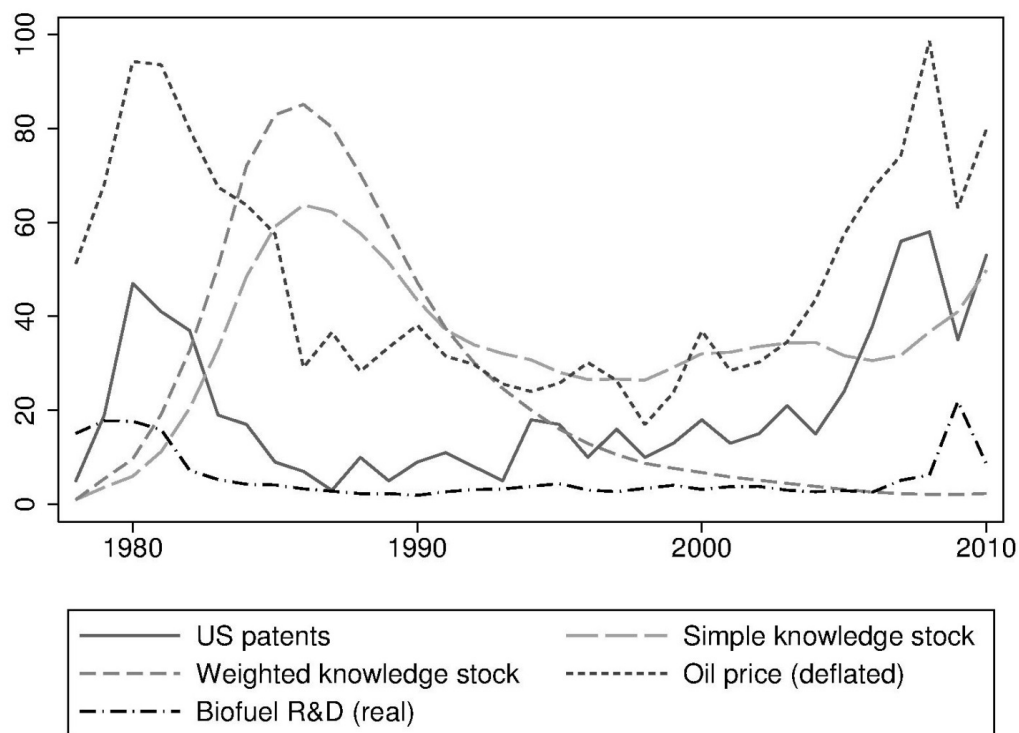


Figure 1-5. The Time Series of Included Explanatory Variables, 1978-2010.



Note:

- (1) Simple and weighted knowledge stocks are normalized using 1978 as the base year.
- (2) Biofuel R&D expenditures (in 10 million \$) are deflated to 2012 dollars.

## Chapter 2

# Evolving Techniques in Production Function Identification Illustrated in the Case of the US Dairy

### 1. Introduction

Two classical issues in the estimation of production function, endogeneity and sample selection, have been revisited owing to the new methodological debates in recent years. Olley and Pakes (1996; hereafter OP), Levinsohn and Petrin (2003; hereafter LP), and Akerberg et al. (2007; hereafter ACF) suggest nonparametric control function approaches to control for unobserved productivity shocks and to correct for simultaneity and selection biases. The most recent contributions to this debate are Wooldridge (2009) and Gandhi et al. (2013; hereafter GNR). Wooldridge (2009) proposes a joint estimation strategy to estimate OP, LP, and ACF more efficiently. GNR argues that the control function approaches suffer from lack of exclusion restrictions and suggests a nonparametric share regression instead of the proxy equation.

There is an extensive body of literature applying the above-mentioned approaches (e.g., Kasahara and Rodrigue 2008; Fan and Firestone 2010; Sieg and Zhang 2012), but no study compares the relative merits of the techniques and none of the applications is in the context of agriculture except Petrick and Kloss (2013). The authors compare OP and LP with the traditional OLS and fixed-effect models and conclude that LP produces more plausible results whereas OP is close to OLS in estimating the production function of the EU agriculture. In this study, focusing on the continuously evolving US dairy sector, we

bring the different approaches together and make a meaningful comparison in order to highlight the importance of properly controlling simultaneity and sample selection in production function estimation.

The US dairy production is characterized by remarkably different patterns of structural changes across traditional and emerging producing regions over the last few decades. Dairy production shifted towards Western and Southern States away from traditional dairy states in the Lake States and Northeast. Production growth in California, Idaho, and New Mexico is primarily responsible for the westward shift (Gould 2010). California replaced Wisconsin as the largest dairy producing state in 1994. Despite the general trend of increasing farm size, small and family-oriented farms still dominate in traditional dairy regions. In contrast, milk production in emerging regions is dominated by large, industrialized dairy farms (Wolf 2003; MacDonald et al. 2007). Underlying the structural changes is productivity in dairy production. On one hand, large dairy farms in emerging regions have adopted industrialized production practices that generate efficiency gains and exploited scale economies by producing high-volume homogeneous output. It provides a strategic cost advantage and a greater likelihood to stay in business with relatively low milk prices and/or high feed costs. On the other hand, integrated small- and medium-farms in traditional dairy regions have certain advantages, which include, for example, favorable agro-climatic conditions for feed production and availability of homegrown feed and family labor. These enable the traditional dairy farms to survive unfavorable economic conditions and to compete with industrialized farms that rely on purchased feed and hired labor.

Employing a unique dataset of the US dairy farms over 1987-2007, we estimate the production technologies of the major US dairy producing regions. Simultaneity and sample selection are carefully corrected following the recent development in the literature, especially Gandhi et al. (2013). The commonly applied control function approaches are found misleading in correcting for biases of coefficient estimates on flexible and quasi-fixed input variables. The results indicate that dairy production in the Pacific region has the highest productivity, followed by Northeast, Mountain, Lake States, and Corn Belt. We decompose the aggregate productivity growth into the contributions of surviving, entering, and exiting farms and find that surviving farms contribute the most. Among the driving factors at the farm and regional level, farm size variation, product diversification, off-farm work, and operator's age are negatively associated with dairy farm productivity, while herd size and vertical integration of individual farms have positive impacts.

The remainder of paper is organized as follows. Section 2 introduces the empirical framework and estimation method. We describe the data and summary statistics in Section 3. Section 4 presents and discusses the estimation results followed by the concluding remarks in Section 5.

## **2. Empirical Framework**

We start with a discussion of the empirical methods to obtain the consistent estimates of production function parameters. A well-known econometric issue is that production decision is typically influenced by factors that are known to the producer but not observed by the econometrician and hence the observed inputs are correlated with the unobservables. This induces endogeneity and, consequently, the OLS estimates are

biased.<sup>16</sup> Standard approaches to correct endogeneity include the fixed effect, first differencing, and instrumental variable (IV) methods, which are usually based on strong assumptions, or have been proven to be unsatisfactory in practice. The fixed effect and first differencing methods require strict exogeneity of inputs conditional on time-invariant farm heterogeneity, which rules out potential responses of input demand to current or past productivity shocks. The IV approach relaxes this assumption and instead requires instruments to be correlated with input demand but not with the unobservables. One natural IV choice is the input price with the assumption of competitive input markets. However, if the assumption holds, the input price unlikely varies across farms. Even if the input price differs across farms, it may not be exogenous to farm decisions. Moreover, farm-level input price is typically not available.

OP suggests a novel control function approach using the firm-level investment as a proxy for the unobservable productivity shock. The proxy function is non-parametrically estimated. OP also incorporates nonrandom exit conditioned on productivity threshold. LP's approach is similar. Instead of investment, LP uses the material input demand as the proxy, which is considered to be more reliable when investment is lumpy and data series contains many zeros or missing values. The control function approach places several structural assumptions on the shape of economic primitives. First, the proxy variable, investment in OP and material input in LP, should be strictly monotonic in the productivity, i.e., more productive farms are expected to have higher capital investment or material input demand. Second, the productivity is assumed

---

<sup>16</sup> See a more detailed discussion in, e.g., Arellano and Bover (1995), Blundell and Bond (2000), and Akerberg et al. (2006).

to be the only unobservable determinant in production function and evolves following an exogenous first-order Markov process. The third assumption is on the timing of input decisions. In both OP and LP, capital is a quasi-fixed input and determined prior to the current productivity shock. To be identifiable, labor is assumed to be a flexible input, the demand of which is made after the realization of the productivity shock. Material is also regarded as a flexible input in LP so as to be used as the proxy.

ACF highlights a potential identification problem associated with the third assumption. ACF argues that if a flexible input, e.g., labor, is also a deterministic function of the unobservable productivity and quasi-fixed inputs, its coefficient is non-parametrically unidentifiable. ACF resolves the issue by identifying all the input coefficients together in the second stage using additional moment conditions. Wooldridge (2009) proposes a more efficient estimation strategy robust to ACF criticism by combining the two estimation steps into one.

GNR is the most recent development in the literature. GNR argues that the control function approaches still suffer from the lack of exclusion restriction and thus create identification problems. GNR suggests to derive additional information from the first-order condition of the profit maximization in order to identify the flexible input coefficient. GNR relies on a nonparametric generalization of the flexible input share regression rather than the proxy function.

In this study, using material demand as the flexible input to correct both simultaneity and attrition biases, we employ the method suggested by GNR to back out the time-varying farm-level productivity. The estimated production function parameters are compared with those obtained from the control function approaches (i.e., LP, ACF

and Wooldridge). Following the insight of OP, exit is conditioned on the productivity threshold that is not observed by the econometrician. Accounting for farm exit is important in our case because we observe significant entry and exit of dairy farms over the sample period. Comparing the different approaches side-by-side clarifies our understanding of farm production technologies and highlights the importance of isolating the confounding effects of simultaneity and sample selection to obtain the consistent estimates. The productivity estimates are then decomposed into the contributions of entering, surviving and existing farms, and used to examine the impact of farm and regional factors on the productivity growth.

We now turn to a brief discussion of the estimation procedure. Assume that to produce  $Q$  units of milk, a farmer combines the inputs of labor ( $L$ ), material ( $M$ ), capital ( $K$ ), feed ( $F$ ), and cows ( $C$ ). Labor includes both full- and part-time hired workers. Material corresponds to the inputs that are renewed on a regular basis, mainly fuel and electricity. Capital refers to the stock of machinery and equipment. Feed includes total amount of purchased feed, and cows are heads of milk cows active in milk production. Milk production also depends on the farm-specific productivity,  $\omega$ , a scalar Hicks neutral productivity parameter, which is known to the farmer but unobservable to the econometrician. We denote the measurement error and idiosyncratic shock by  $u$ . The production technology for farm  $i$  at period  $t$  is:

$$(1) \quad Q_{it} = Y(L_{it}, M_{it}, K_{it}, F_{it}, C_{it}) \exp(\omega_{it} + u_{it}).$$

Taking the logarithm of both sides and using lower case letters to denote logged variables, we have:

$$(2) \quad q_{it} = y(L_{it}, M_{it}, K_{it}, F_{it}, C_{it}) + \omega_{it} + u_{it}.$$

Let  $\mathbf{b} \equiv \{b_l, b_m, b_k, b_f, b_c\}$  denotes the vector of parameters to estimate. For now, we restrict our attention to the Cobb-Douglas production function and rewrite eqn. (2) as:

$$(3) \quad q_{it} = b_l l_{it} + b_m m_{it} + b_k k_{it} + b_f f_{it} + b_c c_{it} + \omega_{it} + u_{it}.$$

We start with the LP estimation. The idea is that while the inputs of  $l_{it}$  and  $f_{it}$  are flexibly determined after  $\omega_{it}$  is observed in period  $t$ , demand for  $m_{it}$  is correlated with  $\omega_{it}$  as well as the quasi-fixed inputs  $k_{it}$  and  $c_{it}$ , the decisions of which are made before the productivity shock. This can be specified as  $m_{it} = m_t(c_{it}, k_{it}, \omega_{it})$ . Assume that the material input demand  $m_{it}$  is strictly monotone in  $\omega_{it}$  for any  $(c_{it}, k_{it})$ . The inverse,  $\omega_{it} \equiv m_t^{-1} = \omega_t(m_{it}, c_{it}, k_{it})$ , is then a valid proxy for productivity in estimating eqn. (3).

Following the insight of OP, we define an indicator function to control sample selection:

$$X_{it} = \begin{cases} 0 & \text{if } \omega_{it} \geq \underline{\omega}_{it}(m_{it}, c_{it}, k_{it}), \\ 1 & \text{otherwise.} \end{cases}$$

where  $X_{it}$  is equal to one if firm  $i$  exits between the periods  $t$  and  $t+1$ , and zero otherwise. We assume that firm  $i$  survives if its productivity ( $\omega_{it}$ ) is at least as high as the threshold ( $\underline{\omega}_{it}$ ).

LP estimates eqn. (3) in three steps:

*Step 1:* Define  $\phi_t(m_{it}, c_{it}, k_{it}) \equiv b_m m_{it} + b_k k_{it} + b_c c_{it} + \omega_t(m_{it}, c_{it}, k_{it})$  and estimate:

$$q_{it} = b_l l_{it} + b_f f_{it} + \phi_t(m_{it}, c_{it}, k_{it}) + u_{it}.$$

Note that all the terms associated with the correlated inputs of  $(m_{it}, c_{it}, k_{it})$  are combined into the flexible function of  $\phi_t$ . The OLS estimation of *Step 1* where  $\phi_t$  is

approximated by a third-order polynomial in  $(m_{it}, c_{it}, k_{it})$  provides the coefficient estimates on the variable inputs,  $b_l$  and  $b_f$ , as well as the estimates of  $\phi_t$ .

*Step 2:* Estimate the exit probability defined in the indicator function above:

$$\begin{aligned} \Pr(X_{it} = 1 | \underline{\omega}_{it}(m_{it}, c_{it}, k_{it})) &= \Pr(\omega_{it} < \underline{\omega}_{it}(m_{it}, c_{it}, k_{it}) | \underline{\omega}_{it}(m_{it}, c_{it}, k_{it})) \\ &= \psi(m_{it}, c_{it}, k_{it}) \equiv P_{it}. \end{aligned}$$

Similarly,  $\psi(\cdot)$  is approximated by a third-order polynomial in  $(m_{it}, c_{it}, k_{it})$  and  $P_{it}$  is predicted from a Probit model.

*Step 3:* Estimate all other coefficients given the law of motion of the productivity, which is assumed to follow a first-order Markov process,

$$\omega_{it} = E[\omega_{it} | \omega_{it-1}, X_{it} = 1] + \xi_{it} = g_t(\omega_{it-1}, P_{it}) + \xi_{it}.$$

where  $\xi_{it}$  indicates the residual of the current productivity innovation. Here the productivity solely depends on one-period lagged productivity and exit probability obtained in *Step 2*. The Generalized Method of Moments (GMM) is used to estimate  $b_m$ ,  $b_k$  and  $b_c$  by relying on the following moment conditions:

$$(4) \quad E[\xi_{it} | m_{it-1}, k_{it}, c_{it}] = 0.$$

ACF highlights a potential collinearity problem in estimating variable input coefficients in *Step 1*. If the variable inputs,  $l_{it}$  and  $f_{it}$ , are also deterministic functions of the unobserved productivity and other quasi-fixed variables, the corresponding coefficients are not identifiable. To solve this issue, ACF modifies *Step 1* by collecting all inputs in  $\phi_t(\cdot)$ , i.e.,  $q_{it} = \phi_t(l_{it}, f_{it}, m_{it}, c_{it}, k_{it}) + u_{it}$ . Moment conditions in eqn. (4) are then combined with the extra moments in eqn. (5) to identify all the parameters in *Step 3*.

$$(5) \quad E[\xi_{it} | l_{it-1}, f_{it-1}] = 0.$$

However, in some cases it is reasonable to assume that demand for labor at period  $t$  is chosen beforehand. For example, labor may have adjustment frictions as the decision involves firing, hiring, and training costs. This is plausible in dairy production as farm labor is usually immobile in the short run due to specific skills and is conditional upon the life-cycle consideration (Petrick and Kloss 2013). If labor is considered to be quasi-fixed or flexible but subject to the adjustment frictions, we can use the moment conditions in eqn. (6) where labor is exogenous to the current productivity innovation. Applying eqn. (6) is likely to generate more efficient estimates than eqn. (5) as the current labor demand links directly to the current output.

$$(6) \quad E[\xi_{it} | l_{it}, f_{it-1}] = 0.$$

Wooldridge (2009) proposes an alternative estimation method that more efficiently estimates the control function approaches and is robust to the collinearity problem highlighted by ACF. The approach jointly estimates a two-equation system with a different set of instruments across the equations, which is more efficient as it utilizes the cross-equation information. Applying Wooldridge approach to the ACF setting, we have

$$(7) \quad q_{it} = \begin{cases} \phi_t(l_{it}, f_{it}, m_{it}, c_{it}, k_{it}) + u_{it} \\ b_l l_{it} + b_m m_{it} + b_k k_{it} + b_f f_{it} + b_c c_{it} + h(l_{it-1}, f_{it-1}, m_{it-1}, c_{it-1}, k_{it-1}, P_{it}) + v_{it} \end{cases}$$

where  $h(l_{it-1}, f_{it-1}, m_{it-1}, c_{it-1}, k_{it-1}, P_{it}) \equiv g_t(\omega_{it-1}(l_{it-1}, f_{it-1}, m_{it-1}, c_{it-1}, k_{it-1}), P_{it})$  and

$v_{it} \equiv \xi_{it} + u_{it}$ . The two equations in eqn. (7) are jointly estimated using different moment conditions across the equations. Specifically, given that  $u_{it}$  is i.i.d., the variables act as their own instruments in the first equation. That is, the moment conditions

$E[u_{it} | l_{it}, f_{it}, m_{it}, c_{it}, k_{it}, \Theta] = 0$  are used where  $\Theta$  is a vector containing all the interaction terms in  $\phi(\cdot)$ . In the second equation,  $m_{it-1}$  is the instrument for  $m_{it}$  as the current material demand is correlated with the productivity innovation,  $\xi_{it}$ . Hence,

$E[v_{it} | l_{it}, f_{it-1}, m_{it-1}, c_{it}, k_{it}, \Lambda] = 0$  are the moment conditions for the second equation, where  $\Lambda$  is a vector containing all the terms in  $h(\cdot)$ .

The most recent development in GNR points out that the control function approaches still suffer from lack of exclusion restrictions. The intuition is that we rely on only five exogenous variables  $(l_{it}, f_{it-1}, m_{it-1}, c_{it}, k_{it})$  to identify the seven coefficients of the variables  $(l_{it}, f_{it}, f_{it-1}, m_{it}, m_{it-1}, c_{it}, k_{it})$  in the second equation in eqn. (7). GNR proposes to identify the coefficient of  $m_{it}$  using additional information from the first-order condition of the profit maximization. Transforming the first-order condition to a nonparametric share regression provides additional information to identify the output elasticity of the flexible input. Elasticities of other input variables are estimated in a similar way as in the control function approaches.

Given the production function in (1), the first order condition with respect to  $m_{it}$  yields,

$$(8) \quad p_t^M = p_t^Q Y_M(L_{it}, M_{it}, K_{it}, F_{it}, C_{it}) \exp(\omega_{it} + u_{it})$$

where  $p_t^M$  and  $p_t^Q$  denote material and output price, respectively, and  $Y_M(\cdot)$  the partial derivative of  $Y(\cdot)$  with respect to  $m_{it}$ . Taking logs of eqn. (8) and differencing it with eqn. (2), we have the share regression as

$$s_{it} = \ln(G(L_{it}, M_{it}, K_{it}, F_{it}, C_{it}) \exp(u_{it})) - u_{it}$$

where  $s_{it} \equiv \log(p_t^M M_{it} / p_t^Q Q_{it})$  denotes the log of the share of material and

$G(\cdot) \equiv Y_M(\cdot) M_{it} / Y(\cdot)$  is the output elasticity of material. Thus, we can recover the idiosyncratic shock  $u_{it}$  and the elasticity  $G(\cdot)$  from the nonparametric share regression.

To identify the production function as a whole, we use the fact that

$$(9) \quad \int \frac{G(L_{it}, M_{it}, K_{it}, F_{it}, C_{it})}{M_{it}} dM_{it} = y(L_{it}, M_{it}, K_{it}, F_{it}, C_{it}) + B(L_{it}, K_{it}, F_{it}, C_{it})$$

where  $B(\cdot)$  is the constant of integration that depends on other inputs. The constant  $B$  and the elasticities of other inputs are identified in a similar way as in the control function approaches by assuming that the productivity follows a first-order Markov process.

Differencing eqns. (2) and (9), we have

$$(10) \quad q_{it} - \int \frac{G(L_{it}, M_{it}, K_{it}, F_{it}, C_{it})}{M_{it}} dM_{it} - u_{it} = B(L_{it}, K_{it}, F_{it}, C_{it}) + \omega_{it}$$

where the estimates for the left-hand side  $\Upsilon \equiv q_{it} - \int \frac{G(L_{it}, M_{it}, K_{it}, F_{it}, C_{it})}{M_{it}} dM_{it} - u_{it}$  is

obtained from the share regression in the first stage. Using eqn. (10), the productivity process can be rewritten as

$$\Upsilon_{it} - B(L_{it}, K_{it}, F_{it}, C_{it}) = g_t(\Upsilon_{it-1} - B(L_{it-1}, K_{it-1}, F_{it-1}, C_{it-1}), P_{it}) + \xi_{it}.$$

The moment conditions  $E[\xi_{it} | L_{it}, K_{it}, F_{it-1}, C_{it}] = 0$  are applied to identify  $B(\cdot)$  and  $\omega_{it}$ .

### 3. Data

We put together a panel dataset of the US dairy farms from the Census of Agriculture over 1987-2007. The Census of Agriculture is the only source of comprehensive agricultural data over time and across the whole country. The Census is conducted every

five years by the US Department of Agriculture (USDA) and surveys all existing farms that produce or normally would have produced \$1,000 or more of agricultural products during the census year. A numeric farm identifier is used to link farms longitudinally. The constructed dataset consists of farm-level observations in the years of 1987, 1992, 1997, 2002, and 2007. For the production function estimation, farms that generate 50% or more of revenue from dairy products (mainly milk) are considered as dairy farms.<sup>17</sup> As the estimation requires detailed information on input expenditures, the observations are restricted to the farms that fill in the long form of the survey, which reports detailed input expenditure data.<sup>18</sup>

For the sampled dairy farms, in each census year we observe (i) revenue from individual crop and livestock products including dairy, (ii) market value of machinery and equipment, (iii) expenditures on feed, contract and hired labor, utilities, fuel, and other livestock related expenses, and (iv) number of dairy cows. Farm-level output is measured by dairy revenue deflated by state-level average milk price. To compute real labor input, we deflate expenditures on contract and hired labor by region-specific hourly wage rates for livestock workers, which are obtained from the Farm Labor Survey by the National Agricultural Statistics Service (NASS) of USDA. Feed costs are the sum of

---

<sup>17</sup> For the descriptive analysis of exits and estimation of exit probability, we utilize the whole sample of all US dairy farms, which is defined as farms producing positive amount of dairy products (not necessarily 50% or higher of revenue from dairy products).

<sup>18</sup> USDA employs a stratified sampling method to select farms to answer the long form questionnaire that contains additional questions on input expenditures. Generally, all farms that are expected to have high revenues are selected, and other farms are also systematically sampled. For example, in the 1992 census, farms in counties with 100 to 199 farms in 1987 were selected at a rate of 1 in 2, and farms in counties with 200 or more farms in 1987 were selected at a rate of 1 in 6 to receive the long form questionnaire.

expenditures on purchased feed and livestock replacement expenses. The corresponding quantity measure is computed by deflating the feed expenditures by 16% protein dairy feed prices at the regional level available from the US Agricultural Prices Summary.<sup>19</sup> Livestock replacement expenses are deflated by state-level prices of milk cow sold for dairy herd replacement. We combine farm expenditures on fuel and electricity to measure the material input, which are deflated by regional diesel and state electricity prices, respectively. Similarly, market value of farm machinery and equipment is deflated by state-specific price index of farm machinery and equipment maintained by the USDA productivity accounts. As the census only reports aggregate input expenditures of all farm outputs, we multiply total expenditure with revenue share of dairy to calculate input expenditure for milk production.

Following the production regions defined by USDA (Blayney 2002), we aggregate 48 continental US states into 10 production regions presented in Figure 1. About 40% of milk is produced in the traditional dairy regions (Lake States, Corn Belt, and Northeast) and the emerging dairy regions (Pacific and Mountain) produce another 40%. Therefore we focus on these five major dairy producing regions in this study. Table 1 presents the descriptive statistics of average input and output measures for each region and the country as a whole. The average farm size, measured in both dairy output and herd size (number of milking cows), is the largest in the Pacific region, followed by the Mountain that includes the emerging dairy states such as Idaho and New Mexico. Farms in the two regions produce on average about six times more milk with an average herd

---

<sup>19</sup> The USDA Agricultural Prices Summary is available on <http://usda.mannlib.cornell.edu/MannUsda/viewDocumentInfo.do?documentID=1003>.

size five times greater than those in traditional dairy regions. A significant number of small farms are located in the traditional dairy states with an average herd size of 150 cows, which is even smaller than the national average. Average herd size in the Corn Belt is slightly lower than that in the Lake States and Northeast.

Table 2 reports summary statistics of major inputs and milk production of average dairy farm in the five regions over the sample period. Average milk production has more than tripled in 2002 compared with that in 1987. Average herd size increased significantly over the same period. However, both output and herd size declined in 2007 back to the level of 1997. Milk production per cow increased from 147 cwt in 1987 to 201 cwt in 2002, and then slightly decreased to 188 cwt in 2007.<sup>20</sup> Table 3 shows summary statistics of farms averaged over time in the five regions grouped by entering, exiting, and surviving farms. A farm is regarded as a new entry in period  $t$  if the farm does not show up in the sample until period  $t$ . Similarly, if a farm produces a positive amount of dairy products in period  $t$  but not in  $t+1$ , the farm exits in  $t$ . A farm is considered as a surviving farm in  $t$  if it appears in both  $t$  and  $t+1$ . Table 3 indicates that average output and herd size of new entering farms are greater than exiting and surviving farms. Exiting farms produce more outputs with larger herd sizes than surviving farms but with lower levels of capital.

Table 4 shows entry and exit trends of dairy farms in each region over the sample period. Entry rate was greater than exit rate in all five regions until 1997. The trend was reversed when exit rate was higher than entry rate over 1997-2007 in all regions,

---

<sup>20</sup> The unit for output is cwt (hundredweight or 100 pounds) as dairy revenue in \$ is deflated by milk price in \$/cwt.

suggesting that the number of dairy farms has been declining over the last decade. The Lake States had the highest entry rate before 1997 but the lowest after that. The net exit rate (the difference between exit and entry rates) is the largest in Lake States followed by Northeast from 1997 to 2007. In contrast, Mountain and Pacific show significant entry rate over the last decade.

## 4. Empirical Results

We start with the discussion of the production function estimation and regional patterns of the productivity growth. We then proceed to quantify farm and regional factors driving the productivity evolution.

### 4.1. Estimation of the Production Function

Production functions are estimated individually for the major dairy regions including Pacific, Mountain, Lake States, Corn Belt, and Northeast. For comparison, we also estimate the production function using data aggregated over all of the ten US regions. Table 5 presents the OLS results and the sum of input coefficients. The hypothesis of constant returns to scale (CRTS) is rejected in all the regions. The productivity can be measured by the estimated constant if it is assumed to be constant across farms and over time, i.e.,  $\omega_{it} = \omega$ . The results indicate that the Pacific has the highest productivity followed by Mountain, Northeast, and Corn Belt. The Lake States is the least productive region, which is obviously not consistent with the strong development history of the dairy industry in the region.

It is well known that the OLS tends to overestimate the coefficients on flexible inputs and underestimate those on quasi-fixed inputs for two reasons. First, endogeneity of the flexible inputs generates positive correlations with the unobserved productivity.

Second, as only surviving farms with productivities above the threshold are included in the sample, we expect the attrition bias leads to a negative relation between the quasi-fixed inputs and the unobserved productivity. This is the case when comparing the OLS results with the GNR estimates presented in Table 6. We see that OLS substantially overestimates the coefficients on material and number of cows and underestimates the coefficient on capital. It implies that the material demand is positively correlated with the unobserved productivity. Also, the selection is negatively associated with capital but positively with herd size. This is consistent with the summary statistics in Table 3 where exiting farms have lower level of capital but larger herd size than surviving farms. From Table 6, the CRTS hypothesis is also rejected for all the regions. However, in contrast to the OLS results that increasing returns to scale (IRTS) is supported in all regions, GNR results show that Mountain, Corn Belt, and Northeast exhibit the decreasing returns to scale (DRTS) while Pacific and Lake States show IRTS.

Table 7 compares the control function approaches with GNR using data aggregated over all of the ten production regions.<sup>21</sup> The results indicate that the coefficient on material is still overestimated in LP and ACF, although it moves toward the GNR estimate in Wooldridge specification. The control function approaches are also misleading in that the correction for quasi-fixed variables goes to the opposite direction. Compared with the OLS results, the coefficient on capital is more underestimated and the coefficient on number of cows is more overestimated.

---

<sup>21</sup> The control function estimation results (LP, ACF, and Wooldridge) for individual regions are included in Appendix.

## 4.2. Regional Productivity Decomposition

Based on the estimated coefficients in GNR for individual regions, we calculate the farm-level productivities and also aggregate them into regional productivities. Following the dynamic productivity decomposition method in Melitz and Polanec (2012), we decompose the regional productivity growth to the contributions of the groups of surviving ( $U$ ), entering ( $E$ ) and exiting ( $X$ ) farms. Let  $s_{it}$  and  $s_{Gt}$  denote the market shares of farm  $i$  and group  $G$  in a region at time  $t$ . Regional productivity  $\Omega_t$  is calculated as the weighted sum of group productivities:

$$(11) \quad \Omega_t = \sum_{G=U,E,X} \left( s_{Gt} \sum_{i \in G} \frac{s_{it}}{s_{Gt}} \omega_{it} \right) \equiv \sum_{G=U,E,X} \Omega_{Gt}$$

where  $\Omega_{Gt} \equiv s_{Gt} \sum_{i \in G} \frac{s_{it}}{s_{Gt}} \omega_{it}$  is the contribution of the farms in group  $G$ . Table 8 presents the decomposition results grouped by the census years. The first row in each census year shows the regional aggregate productivity  $\Omega_t$ , followed by the contributions of the individual groups. By definition, no farms entered in the initial period and exited in the last period, so we have only two groups in 1987 and 2007.

Averaged over the sample period, the Pacific exhibits the highest regional productivity, followed by Northeast, Mountain, Lake States, and Corn Belt. Surviving farms are shown to have played the most important role in the regional productivity growth. Entering farms generally contributed more than exiting farms except 1997 when The Federal Agriculture Improvement and Reform Act was implemented. In 1997, the significant policy reform reformulated Federal Milk Marketing Order (FMMO) in a way that the number of milk marketing orders was consolidated from 42 to 11 and minimum

prices received by dairy farmers were recalculated by 2000. This policy change led to a significant transition of the US dairy industry from 1997 to 1999.

We analyze if the policy reform had a considerable effect on exiting farms, by presenting a simple Probit estimation of farm exit with FMMO dummy and other farm- and regional- factors as right-hand side variables. Our estimating equation is:

$$(12) \quad \Pr(Exit_{it} = 1) = \Phi(\beta_1 FMMO_t + \beta_2 THI_{ct} + \beta_3 Var_{st} + X_{it}\gamma).$$

where farm  $i$  is located in county  $c$ , state  $s$ , and FMMO region  $k$ .  $Exit_{it}$  is equal to one if farm  $i$  exited between the period  $t$  and  $t+1$ .  $FMMO_t$  is a dummy variable that equals 1 after 1997. To account for heterogeneous regional agro-climatic conditions, we construct the Temperature Humidity Index ( $THI_{ct}$ ) at the county level.<sup>22</sup> THI has been widely used to explain the combined effect of temperature and humidity on production of livestock and related products. Figure 2 shows the average THI over the sample period. It is expected that dairy cows start experiencing heat stress when the THI exceeds around 70. The Southern Plains, Delta States, and Southeast regions have average THIs in the range of 63 to 72, suggesting that the climate in these regions is not suitable for dairy production.  $Var_{st}$  is the variance of herd size within state  $s$ . We collect farm-level

---

<sup>22</sup> The THI for county  $c$  at time  $t$  is calculated as  $THI_{ct} = T_{ct}^{db} + 0.36T_{ct}^{dp} + 41.5$ , where  $T_{ct}^{db}$  denote the dry-bulb temperature ( $^{\circ}C$ ) and  $T_{ct}^{dp}$  the dew-point temperature ( $^{\circ}C$ ) (Collier and Zimbelman 2007). The temperature data are derived from the Parameter-Elevation Regressions on Independent Slopes Model (PRISM), which provides estimates of precipitation and temperature at 4×4 kilometer grid cells for the entire US and is considered as one of the most reliable climatic data source on a small scale. (See <http://www.prism.oregonstate.edu> for more details). We construct the THI by overlapping a county's boundary on the PRISM data and then taking simple average across grid cells.

controls in the vector  $X_{it}$  including herd size, indices for diversification,<sup>23</sup> vertical integration,<sup>24</sup> off-farm work,<sup>25</sup> and operator's age.

Table 9 provides summary statistics. Average THI varies from 49 to 55, implying that the major dairy regions have favorable climate conditions for dairy production. Farm size varies considerably across the regions. As expected, traditional dairy production regions consist of small farms and have relatively homogeneous farm size distributions compared with emerging regions. Vertical integration measures the magnitude of farm-produced input such as homegrown feed. The corresponding index varies from 0 to 1. The value of 0 means that a farm purchases all feed inputs and 1 indicates a fully integrated farm producing 100% of dairy feed. Vertical integration is significant and more common in traditional dairy regions than emerging regions. In Lake States, the index shows an average of 0.60, while it falls to 0.38 in Pacific region. Diversification also ranges from 0 to 1. It is close to 0 for fully specialized dairy farms and to 1 for fully diversified farms. Dairy farms in Pacific are the most specialized with the index of 0.11, whereas Corn Belt dairy farms are the least specialized with the index of 0.30. Off-farm work ranging from 1.45 to 1.67 across regions implies that days worked off farm are less than 50 days on average.

---

<sup>23</sup> Farm diversification index is calculated as (total farm revenue – dairy revenue)/total farm revenue.

<sup>24</sup> Vertical integration is defined as (dairy value-added/dairy revenue), where dairy value-added is dairy revenue net of purchased feed, dairy cow replacement costs, and hired labor. See Sumner and Wolf (2002) for more details.

<sup>25</sup> Off-farm work in the Census of Agriculture is measured as a categorical variable ranging from 1 to 5 (1 for 0, 2 for 1-49, 3 for 50-99, 4 for 100-199 and 5 for 200 or more days worked off farm).

Table 10 provides the marginal effects of Probit regression, where standard errors are robust clustered by the county. All variables are statistically significant except THI. Specifically, the results show that FMMO affected more farms to exit. Exit probabilities are higher for farms located in the states with more dispersed farm size distributions. Farms with larger herd sizes and higher vertical integrations are more likely to stay in business. However, farms with higher diversifications have higher probabilities to exit. Off-farm work and operator's age are also positively associated with exit probabilities. Overall, the results indicate that although FMMO is justified for guaranteeing minimum prices and stabilizing price fluctuations, it may have provided incentives for large farms to stay in business and expand farm sizes, while not keeping small farms from exit.

### 4.3. Driving Factors of the Productivity Growth

To explain the changes in farm productivity, we regress the GNR productivity estimates on farm and regional variables with the baseline model specified as:

$$(13) \quad \omega_{it} = \beta_0 + \beta_1 FMMO_t + \beta_2 THI_{ct} + \beta_3 Var_{st} + X_{it} \gamma + \alpha_{sf} + \varepsilon_{it}.$$

where  $\omega_{it}$  is the productivity of farm  $i$  located in county  $c$ , state  $s$ , and FMMO region  $k$ . Eqn. (13) also includes the state-FMMO fixed effect to control for persistent productivity differences that may exist between states and FMMO regions. Thus the coefficients are identified based on the differences within state-FMMO areas. All the variables are log transformed so that the coefficients are interpreted as elasticities.

Table 11 presents the results with the robust standard errors clustered by the county. Without including fixed effects (Column I), we find positive effects of FMMO reform and herd size but negative effects of THI, farm size variance, vertical integration, diversification, and off-farm work on farm productivity. However, the impact of vertical

integration becomes positive when the fixed effects are included (Columns II-VI). With different sets of fixed effects, we find almost identical effects of farm size variance, herd size in farm, diversification, and off-farm work. For example, in the baseline specification with state-FMMO fixed effects (Column V), the results imply that a one percent increase in farm size variance reduces farm productivity by 0.03 percent. At the farm-level, a one percent increase in herd size and vertical integration increases productivity by 0.06 and 0.01 percent, respectively. Diversification, off-farm work, and operator's age are all negatively associated with productivity, with elasticity of 0.02, 0.03, and 0.02, respectively.

## **5. Conclusion**

Joining the evolving literature, this study makes a side-by-side comparison of the recently developed techniques for production function estimation and highlights the importance of carefully controlling simultaneity and sample selection. Employing a unique panel dataset on the US dairy farms, we apply the estimation techniques to regional milk production and disclose the driving factors of regional productivity differences and changes. The sample selection issue is accounted for not only in the estimation procedure but also in the productivity decomposition.

The results show that the Pacific region has the highest productivity, followed by Northeast, Mountain, Lake States, and Corn Belt. Decomposition of aggregate regional productivities indicates that surviving farms contributed more to the regional productivity growth than entering and exiting farms. At the state level, variation of herd size is negatively associated with farm productivity. At the farm level, herd size and vertical

integration have positive impact on the productivity, while the effects of diversification, off-farm work, and operator's age are negative.

We conclude with several caveats. First, the quantity measures of production inputs are constructed under data limitation where we assume that input expenditures are proportional to farm revenue. Second, homegrown feed and family labor are not taken into account. Such inputs are not reported in the census. The cost saving effects could be important especially for farms in the traditional dairy producing regions. By limiting the samples to the farms generating at least 50% revenue from dairy, we may partially alleviate the concern for homegrown feed. Finally, the applied empirical framework is essentially static. A potential extension could be to develop a model of dynamic industrial development to examine the inter-temporal aspects of the productivity growth and the evolution of industry.

Table 2-1. Descriptive Statistics of Production Regions: Sample Mean and Standard Deviation (in the parentheses).

	Output (cwt)	Labor (hour)	Capital <sup>1</sup>	Material <sup>2</sup>	Feed <sup>3</sup>	# of Cows	# of Obs.
National Average	49,243 (119,659)	9,411 (24,324)	174,492 (581,215)	1,062 (2,675)	1,112 (3,187)	268 (603)	47,164
1. Pacific	166,183 (218,820)	25,460 (39,295)	137,713 (265,116)	2,490 (3,714)	4,382 (6,368)	840 (1,087)	4,343
2. Mountain	148,337 (260,497)	27,810 (50,571)	541,013 (1,160,325)	3,300 (7,315)	3,385 (6,462)	802 (1,334)	2,665
3. Lake States	28,157 (54,995)	5,453 (13,467)	69,204 (97,600)	713 (1,214)	437 (1,105)	147 (247)	13,898
4. Corn Belt	25,833 (84,240)	4,515 (15,718)	61,461 (145,042)	648 (2,063)	489 (2,096)	145 (368)	5,908
5. Northeast	28,605 (49,211)	7,017 (15,872)	366,231 (996,534)	662 (1,051)	582 (1,127)	155 (227)	10,552

Note:

- (1) <sup>1</sup> Market values of farm machinery and equipment are deflated by state-specific price index from USDA productivity accounts (Alabama in 1996 = 1).
- (2) <sup>2</sup> Fuel and electricity expenditures are deflated by diesel (\$/gallon) and electricity (\$/million Btu) prices, respectively.
- (3) <sup>3</sup> Purchased feed and livestock expenses are deflated by 16% protein dairy feed (\$/ton) and milk cow replacement (\$/head) costs, respectively.

Table 2-2. Descriptive Statistics of Major Inputs and Output: Sample Mean and Standard Deviation (in the parentheses).

	1987	1992	1997	2002	2007
Output	22,886	36,288	48,195	72,064	45,383
(cwt)	(42,342)	(68,687)	(90,217)	(142,621)	(126,524)
Labor	6,823	9,003	10,510	12,357	8,300
(hours)	(14,875)	(19,437)	(22,034)	(25,800)	(25,015)
Capital <sup>1</sup>	77,041	109,010	174,121	225,931	193,356
	(439,658)	(249,652)	(671,963)	(694,236)	(651,702)
Material <sup>2</sup>	465	687	907	1,386	1,002
	(713)	(1,094)	(1,814)	(2,629)	(3,100)
Feed <sup>3</sup>	575	925	1,112	1,421	1,071
	(1705)	(2,195)	(2,685)	(3,446)	(2,394)
# Cows	155	222	303	358	241
	(248)	(373)	(541)	(685)	(633)
# Obs.	1,562	1,900	1,978	11,786	20,141

Note:

- (1) <sup>1</sup> Market values of farm machinery and equipment are deflated by state-specific price index from USDA productivity accounts (Alabama in 1996 = 1).
- (2) <sup>2</sup> Fuel and electricity expenditures are deflated by diesel (\$/gallon) and electricity (\$/million Btu) prices, respectively.
- (3) <sup>3</sup> Purchased feed and livestock expenses are deflated by 16% protein dairy feed (\$/ton) and milk cow replacement (\$/head) costs, respectively.

Table 2-3. Descriptive Statistics of Entering, Exiting, and Surviving Dairy Farms  
(Averaged over 1987-2007).

	All	Entering	Exiting	Surviving
Output (cwt)	49,243 (119,659)	61,949 (152,646)	52,688 (102,481)	50,398 (122,507)
Labor (hours)	9,411 (24,324)	10,828 (27,411)	9,777 (20,508)	9,404 (24,510)
Capital <sup>1</sup>	174,492 (581,215)	167,831 (457,380)	180,358 (583,718)	202,886 (696,796)
Material <sup>2</sup>	1,062 (2,675)	1,201 (3,096)	1,052 (1,957)	1,059 (2,806)
Feed <sup>3</sup>	1,112 (3,187)	1,403 (3,988)	1,092 (2519)	1,104 (3,242)
# of Cows	268 (603)	323 (740)	284 (528)	265 (611)
# Obs.	47,164	8,033	5,635	25,414

Note:

- (1) <sup>1</sup> Market values of farm machinery and equipment are deflated by state-specific price index from USDA productivity accounts (Alabama in 1996 = 1).
- (2) <sup>2</sup> Fuel and electricity expenditures are deflated by diesel (\$/gallon) and electricity (\$/million Btu) prices, respectively.
- (3) <sup>3</sup> Purchased feed and livestock expenses are deflated by 16% protein dairy feed (\$/ton) and milk cow replacement (\$/head) costs, respectively.

Table 2-4. Descriptive Statistics of Entry and Exit Rate by Production Region.

	Entry (%)				Exit (%)			
	1987- 1992	1992- 1997	1997- 2002	2002- 2007	1987- 1992	1992- 1997	1997- 2002	2002- 2007
National Average	29.8	30.5	29.8	30.4	20.0	24.9	46.6	47.2
1. Pacific	29.9	29.2	38.3	36.0	22.2	23.1	40.3	45.1
2. Mountain	25.7	27.4	38.5	48.5	16.9	22.4	41.5	50.6
3. Lake States	35.2	36.3	24.1	19.2	23.5	31.6	48.8	42.3
4. Corn Belt	24.5	24.7	33.4	34.0	13.8	16.5	42.2	48.8
5. Northeast	26.9	27.7	25.9	27.8	19.5	23.5	46.0	43.9

Table 2-5. The OLS Regression Results.

	All Regions	Pacific	Mountain	Lake States	Corn Belt	Northeast
Labor	0.05 <sup>c</sup> (0.001)	0.10 <sup>c</sup> (0.005)	0.06 <sup>c</sup> (0.006)	0.05 <sup>c</sup> (0.001)	0.04 <sup>c</sup> (0.003)	0.03 <sup>c</sup> (0.002)
Feed	0.10 <sup>c</sup> (0.001)	0.07 <sup>c</sup> (0.005)	0.10 <sup>c</sup> (0.006)	0.10 <sup>c</sup> (0.002)	0.09 <sup>c</sup> (0.004)	0.12 <sup>c</sup> (0.003)
Capital	0.06 <sup>c</sup> (0.001)	0.04 <sup>c</sup> (0.005)	0.03 <sup>c</sup> (0.006)	0.07 <sup>c</sup> (0.003)	0.08 <sup>c</sup> (0.004)	0.06 <sup>c</sup> (0.003)
Material	0.08 <sup>c</sup> (0.002)	0.06 <sup>c</sup> (0.006)	0.06 <sup>c</sup> (0.007)	0.13 <sup>c</sup> (0.004)	0.08 <sup>c</sup> (0.005)	0.06 <sup>c</sup> (0.004)
# Cows	0.74 <sup>c</sup> (0.002)	0.74 <sup>c</sup> (0.008)	0.75 <sup>c</sup> (0.01)	0.75 <sup>c</sup> (0.005)	0.75 <sup>c</sup> (0.007)	0.78 <sup>c</sup> (0.006)
Constant	3.85 <sup>c</sup> (0.01)	4.12 <sup>c</sup> (0.04)	4.03 <sup>c</sup> (0.05)	3.63 <sup>c</sup> (0.02)	3.85 <sup>c</sup> (0.03)	3.88 <sup>c</sup> (0.03)
Sum	1.03	1.01	1.00	1.1	1.04	1.05
CRTS <sup>1</sup>	reject	reject	reject	reject	reject	reject
Year Fixed Effects	yes	yes	yes	yes	yes	yes
State Fixed Effects	yes	yes	yes	yes	yes	yes
# Obs.	47,164	4,343	2,665	13,898	5,908	10,552

Note:

(1) <sup>1</sup> Fail to reject at the null hypothesis at the 5% significance level.

(2) Standard errors are in the parentheses.

(3) The superscripts, a, b, and c, denote significance at 0.10, 0.05, and 0.01 levels, respectively.

Table 2-6. GNR Estimation Results.

	All Regions	Pacific	Mountain	Lake States	Corn Belt	Northeast
Labor	0.05 <sup>c</sup> (0.00)	0.09 <sup>c</sup> (0.00)	0.07 <sup>c</sup> (0.00)	0.08 <sup>c</sup> (0.00)	0.01 <sup>c</sup> (0.00)	0.02 <sup>c</sup> (0.00)
Feed	0.11 <sup>c</sup> (0.00)	0.14 <sup>c</sup> (0.00)	0.09 <sup>c</sup> (0.00)	0.09 <sup>c</sup> (0.00)	0.10 <sup>c</sup> (0.00)	0.18 <sup>c</sup> (0.00)
Capital	0.06 <sup>c</sup> (0.00)	0.12 <sup>c</sup> (0.00)	0.05 <sup>c</sup> (0.00)	0.10 <sup>c</sup> (0.00)	0.12 <sup>c</sup> (0.00)	0.005 <sup>c</sup> (0.00)
Material	0.04 <sup>c</sup> (0.00)	0.04 <sup>c</sup> (0.00)	0.04 <sup>c</sup> (0.00)	0.05 <sup>c</sup> (0.00)	0.04 <sup>c</sup> (0.00)	0.05 <sup>c</sup> (0.00)
# Cows	0.66 <sup>c</sup> (0.00)	0.60 <sup>c</sup> (0.00)	0.69 <sup>c</sup> (0.00)	0.67 <sup>c</sup> (0.00)	0.71 <sup>c</sup> (0.00)	0.71 <sup>c</sup> (0.00)
Sum	0.94 (0.00)	1.01 (0.00)	0.95 (0.00)	1.00 (0.00)	0.99 (0.00)	0.98 (0.00)
CRTS <sup>1</sup>	reject	reject	reject	reject	reject	reject
# Obs.	47,164	4,343	2,665	13,898	5,908	10,552

Note:

(1) <sup>1</sup> Fail to reject the null hypothesis at the 5% significance level.

(2) Standard errors are in the parentheses.

(3) The superscripts, a, b, and c, denote significance at 0.10, 0.05, and 0.01 levels, respectively.

Table 2-7. Control Function Approach Estimation Results in All Regions.

	OLS	GNR	LP	ACF	Wooldridge
Labor	0.05 <sup>c</sup> (0.001)	0.05 <sup>c</sup> (0.00)	0.04 <sup>c</sup> (0.001)	0.05 <sup>c</sup> (0.005)	0.05 <sup>c</sup> (0.00)
Feed	0.10 <sup>c</sup> (0.001)	0.11 <sup>c</sup> (0.00)	0.10 <sup>c</sup> (0.003)	0.07 <sup>c</sup> (0.01)	0.09 <sup>c</sup> (0.00)
Capital	0.06 <sup>c</sup> (0.001)	0.06 <sup>c</sup> (0.00)	0.03 <sup>c</sup> (0.003)	0.03 <sup>c</sup> (0.002)	0.05 <sup>c</sup> (0.00)
Material	0.08 <sup>c</sup> (0.002)	0.04 <sup>c</sup> (0.00)	0.07 <sup>c</sup> (0.01)	0.07 <sup>c</sup> (0.01)	0.05 <sup>c</sup> (0.01)
# Cows	0.74 <sup>c</sup> (0.002)	0.66 <sup>c</sup> (0.00)	0.80 <sup>c</sup> (0.01)	0.81 <sup>c</sup> (0.02)	0.81 <sup>c</sup> (0.00)
Constant	3.85 <sup>c</sup> (0.01)				
Sum	1.03 (0.00)	0.94 (0.00)	1.04 (0.00)	1.03 (0.00)	1.05 (0.00)
CRTS <sup>1</sup>	reject	reject	reject	reject	reject
# Obs.	47,164	47,164	47,164	47,164	47,164

Note:

- (1) <sup>1</sup> Fail to reject the null hypothesis at the 5% significance level.
- (2) Standard errors are in the parentheses.
- (3) The superscripts, a, b, and c, denote significance at 0.10, 0.05, and 0.01 levels, respectively.

Table 2-8. Dynamic Olley-Oakes Productivity Decomposition Results.

	All Regions	Pacific	Mountain	Lake States	Corn Belt	Northeast
<i>1987</i>						
All farms	4.96	4.28	4.00	4.07	3.95	4.21
Surviving farms	3.89	3.32	2.70	3.08	3.44	3.56
Entering farms	-	-	-	-	-	-
Exiting farms	1.06	0.96	1.29	0.99	0.51	0.64
<i>1992</i>						
All farms	5.44	4.56	4.32	4.69	4.28	4.78
Surviving farms	2.90	2.59	2.48	2.19	2.25	2.32
Entering farms	1.41	1.06	0.91	1.31	1.30	1.30
Exiting farms	1.12	0.90	0.92	1.17	0.72	1.15
<i>1997</i>						
All farms	5.58	4.87	4.71	5.17	4.34	4.83
Surviving farms	2.05	1.89	2.06	1.35	1.92	1.81
Entering farms	1.39	1.23	1.17	1.50	0.89	1.10
Exiting farms	2.13	1.74	1.47	2.31	1.52	1.91
<i>2002</i>						
All farms	5.72	5.08	4.95	4.56	4.69	4.81
Surviving farms	2.71	2.24	2.20	2.53	1.92	2.90
Entering farms	1.53	1.51	1.53	1.01	1.65	0.76
Exiting farms	1.47	1.33	1.21	1.01	1.11	1.14
<i>2007</i>						
All farms	4.83	4.37	4.22	4.21	4.05	4.42
Surviving farms	3.77	3.38	3.03	3.49	2.95	3.77
Entering farms	1.06	0.98	1.18	0.71	1.09	0.64
Exiting farms	-	-	-	-	-	-
<i>1987-2007</i>						
All farms average	5.21	4.72	4.51	4.37	4.25	4.57

Table 2-9. Descriptive Statistics of Farm- and Regional- Variables by Production Region.

	Pacific	Mountain	Lake States	Corn Belt	Northeast
$THI_{ct}$	54.91 (4.68)	49.34 (4.64)	49.42 (1.49)	53.71 (2.34)	50.72 (2.91)
$Var_{st}$	1,107,897 (500,046)	1,484,968 (1,331,156)	59,992 (33,119)	134,588 (170,508)	49,585 (37,004)
$X_{it}$					
# Cows	840 (1,087)	802 (1,334)	147 (247)	145 (368)	155 (227)
Vertical integration	0.38 (0.37)	0.37 (0.52)	0.60 (0.30)	0.51 (0.89)	0.50 (0.76)
Diversification	0.11 (0.16)	0.20 (0.25)	0.20 (0.19)	0.30 (0.24)	0.13 (0.16)
Off-farm work	1.67 (1.43)	1.67 (1.38)	1.45 (1.13)	1.59 (1.27)	1.52 (1.20)
Operator's age	53.60 (12.64)	51.64 (12.03)	50.72 (11.14)	51.85 (12.30)	51.84 (12.55)
# Obs.	4,343	2,665	13,898	5,908	10,552

Table 2-10. Marginal Effect in Probit Estimation of Exit.

Dependent variable: $Exit_{it}$	
$FMMO_t$	0.11 <sup>c</sup> (0.01)
$THI_{ct}$	-0.05 (0.07)
$Var_{st}$	0.01 <sup>c</sup> (0.003)
# Cows	-0.05 <sup>c</sup> (0.004)
Vertical integration	-0.04 <sup>c</sup> (0.008)
Diversification	0.01 <sup>b</sup> (0.004)
Off-farm work	0.05 <sup>c</sup> (0.009)
Operator's age	0.08 <sup>c</sup> (0.01)
# Obs.	16,120

Note:

- (1) Robust standard errors clustered by the county are in the parentheses.
- (2) The superscripts, a, b, and c, denote significance at 0.10, 0.05, and 0.01 levels, respectively.

Table 2-11. Determinants of Farm-Level Productivity.

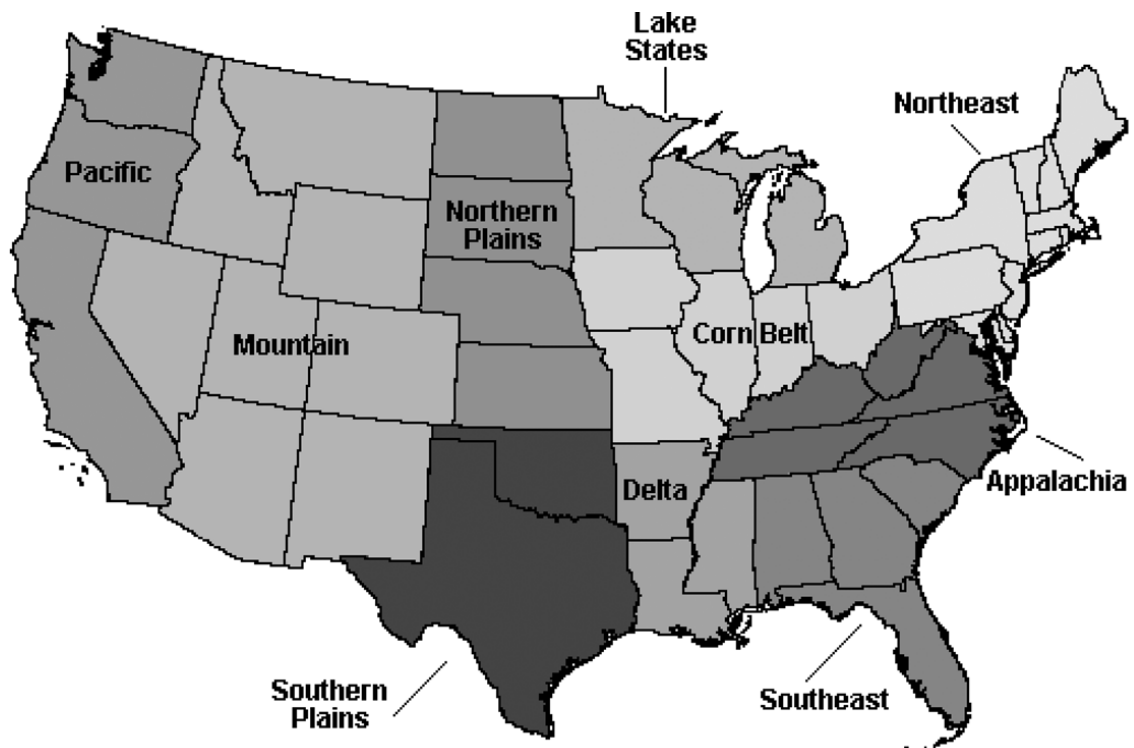
	I	II	III	IV	V	VI
$FMMO_t$	0.06 <sup>a</sup> (0.03)	0.14 <sup>c</sup> (0.02)	0.01 (0.02)	0.14 <sup>c</sup> (0.03)	0.009 (0.02)	
$THI_{ct}$	-0.92 <sup>c</sup> (0.34)	0.25 (0.47)	-0.02 (0.08)	-1.67 <sup>c</sup> (0.23)	-0.02 (0.07)	-0.18 <sup>c</sup> (0.06)
$Var_{st}$	-0.06 <sup>c</sup> (0.01)	-0.07 <sup>c</sup> (0.008)	-0.03 <sup>c</sup> (0.006)	-0.09 <sup>c</sup> (0.01)	-0.03 <sup>c</sup> (0.006)	
$X_{it}$						
# Cows	0.07 <sup>c</sup> (0.01)	0.06 <sup>c</sup> (0.01)	0.06 <sup>c</sup> (0.005)	0.05 <sup>c</sup> (0.009)	0.06 <sup>c</sup> (0.005)	0.04 <sup>c</sup> (0.004)
Vertical integration	-0.08 <sup>c</sup> (0.01)	0.009 <sup>b</sup> (0.004)	0.01 <sup>c</sup> (0.004)	0.01 (0.009)	0.01 <sup>c</sup> (0.004)	0.02 <sup>c</sup> (0.004)
Diversification	-0.07 <sup>c</sup> (0.008)	-0.02 <sup>c</sup> (0.002)	-0.02 <sup>c</sup> (0.002)	-0.03 <sup>c</sup> (0.004)	-0.02 <sup>c</sup> (0.002)	-0.03 <sup>c</sup> (0.002)
Off-farm work	-0.02 <sup>c</sup> (0.007)	-0.03 <sup>c</sup> (0.003)	-0.03 <sup>c</sup> (0.003)	-0.04 <sup>c</sup> (0.004)	-0.03 <sup>c</sup> (0.003)	-0.01 <sup>c</sup> (0.003)
Operator's age	0.01 (0.02)	-0.04 <sup>c</sup> (0.01)	-0.02 <sup>b</sup> (0.01)	0.00004 (0.01)	-0.02 <sup>b</sup> (0.01)	-0.01 (0.009)
Constant	8.48 (1.29)	4.39 <sup>b</sup> (1.89)	4.99 <sup>c</sup> (0.33)	11.97 <sup>c</sup> (0.82)	5.00 <sup>c</sup> (0.31)	5.26 <sup>c</sup> (0.26)
Fixed effects						
County		X				
State			X			
FMMO				X		
State-FMMO					X	
State-FMMO-Year						X
$R^2$	0.09	0.72	0.69	0.57	0.69	0.76
$F$	41.41	103.77	93.98	92.92	91.41	83.70
# Obs.	35,304	35,304	35,304	35,304	35,304	35,304

Note:

(1) Robust standard errors clustered by the county are in the parentheses.

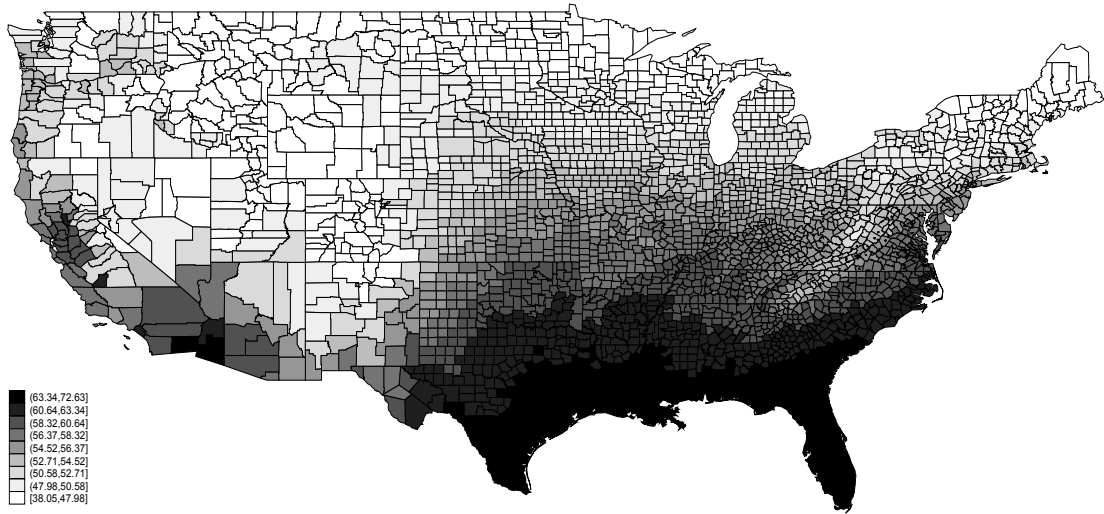
(2) The superscripts, a, b, and c, denote significance at 0.10, 0.05, and 0.01 levels, respectively.

Figure 2-1. Farm Production Regions in the United States.



Source: Mishra, Moss and Erickson (2009), p. 50. Figure 1.

Figure 2-2. Average County-Level Temperature Humidity Index, 1987-2007.



## 6. Appendix

Table 2-A1. LP Estimation Results.

	All Regions	Pacific	Mountain	Lake States	Corn Belt	Northeast
Labor	0.04 <sup>c</sup> (0.001)	0.10 <sup>c</sup> (0.01)	0.06 <sup>c</sup> (0.007)	0.04 <sup>c</sup> (0.002)	0.03 <sup>c</sup> (0.003)	0.02 <sup>c</sup> (0.003)
Feed	0.10 <sup>c</sup> (0.003)	0.06 <sup>c</sup> (0.008)	0.10 <sup>c</sup> (0.009)	0.10 <sup>c</sup> (0.004)	0.07 <sup>c</sup> (0.005)	0.11 <sup>c</sup> (0.007)
Capital	0.03 <sup>c</sup> (0.003)	0.02 <sup>c</sup> (0.01)	0.01 (0.01)	0.02 <sup>c</sup> (0.005)	0.05 <sup>c</sup> (0.005)	0.01 <sup>c</sup> (0.004)
Material	0.07 <sup>c</sup> (0.01)	0.09 <sup>c</sup> (0.03)	0.13 <sup>c</sup> (0.04)	0.05 <sup>b</sup> (0.02)	-0.03 (0.03)	0.11 <sup>c</sup> (0.03)
# Cows	0.80 <sup>c</sup> (0.01)	0.78 <sup>c</sup> (0.02)	0.75 <sup>c</sup> (0.02)	0.83 <sup>c</sup> (0.02)	0.88 <sup>c</sup> (0.03)	0.78 <sup>c</sup> (0.03)
Sum	1.04	1.05	1.05	1.04	1.00	1.03
CRTS <sup>1</sup>	reject	reject	reject	reject	reject	reject
# Obs.	47,164	4,343	2,665	13,898	5,908	10,552

Note:

(1) <sup>1</sup> Fail to reject the null hypothesis at the 5% significance level.

(2) Standard errors are in the parentheses.

(3) The superscripts, a, b, and c, denote significance at 0.10, 0.05, and 0.01 levels, respectively.

Table 2-A2. ACF Estimation Results with  $l_t$  as Instrument.

	All Regions	Pacific	Mountain	Lake States	Corn Belt	Northeast
Labor	0.05 <sup>c</sup> (0.005)	0.03 (0.02)	0.06 <sup>b</sup> (0.02)	0.04 <sup>c</sup> (0.007)	0.06 <sup>c</sup> (0.01)	0.05 <sup>c</sup> (0.01)
Feed	0.07 <sup>c</sup> (0.01)	0.12 <sup>c</sup> (0.03)	0.04 (0.03)	0.14 <sup>c</sup> (0.02)	0.009 (0.03)	0.12 <sup>c</sup> (0.03)
Capital	0.03 <sup>c</sup> (0.002)	0.03 <sup>c</sup> (0.009)	0.02 <sup>b</sup> (0.01)	0.02 <sup>c</sup> (0.007)	0.05 <sup>c</sup> (0.008)	0.01 <sup>c</sup> (0.005)
Material	0.07 <sup>c</sup> (0.01)	0.08 <sup>b</sup> (0.03)	0.09 <sup>b</sup> (0.03)	0.05 <sup>a</sup> (0.02)	0.02 (0.04)	0.11 <sup>c</sup> (0.02)
# Cows	0.81 <sup>c</sup> (0.02)	0.78 <sup>c</sup> (0.04)	0.83 <sup>c</sup> (0.06)	0.79 <sup>c</sup> (0.02)	0.89 <sup>c</sup> (0.05)	0.75 <sup>c</sup> (0.04)
Sum	1.03	1.04	1.04	1.04	1.03	1.04
CRTS <sup>1</sup>	reject	reject	reject	reject	reject	reject
# Obs.	47,164	4,343	2,665	13,898	5,908	10,552

Note:

- (1) <sup>1</sup> Fail to reject the null hypothesis at the 5% significance level.
- (2) Standard errors are in the parentheses.
- (3) The superscripts, a, b, and c, denote significance at 0.10, 0.05, and 0.01 levels, respectively.

Table 2-A3. Wooldridge Estimation Results.

	All Regions	Pacific	Mountain	Lake States	Corn Belt	Northeast
Labor	0.05 <sup>c</sup> (0.00)	0.05 <sup>c</sup> (0.01)	0.08 <sup>c</sup> (0.02)	0.05 <sup>c</sup> (0.00)	0.04 <sup>c</sup> (0.00)	0.05 <sup>c</sup> (0.00)
Feed	0.09 <sup>c</sup> (0.00)	0.08 <sup>c</sup> (0.01)	0.09 <sup>c</sup> (0.01)	0.08 <sup>c</sup> (0.00)	0.09 <sup>c</sup> (0.01)	0.14 <sup>c</sup> (0.01)
Capital	0.05 <sup>c</sup> (0.00)	0.03 <sup>c</sup> (0.00)	0.04 <sup>c</sup> (0.01)	0.07 <sup>c</sup> (0.00)	0.05 <sup>c</sup> (0.01)	0.01 <sup>b</sup> (0.00)
Material	0.05 <sup>c</sup> (0.01)	0.03 (0.02)	0.07 <sup>a</sup> (0.03)	0.05 <sup>b</sup> (0.02)	0.08 <sup>b</sup> (0.03)	0.03 (0.02)
# Cows	0.81 <sup>c</sup> (0.00)	0.82 <sup>c</sup> (0.02)	0.73 <sup>c</sup> (0.03)	0.77 <sup>c</sup> (0.02)	0.80 <sup>c</sup> (0.03)	0.80 <sup>c</sup> (0.02)
Sum	1.05	1.01	1.01	1.02	1.06	1.03
CRTS <sup>1</sup>	reject	not reject	not reject	reject	reject	reject
# Obs.	47,164	4,343	2,665	13,898	5,908	10,552

Note:

(1) <sup>1</sup> Fail to reject the null hypothesis at the 5% significance level.

(2) Standard errors are in the parentheses.

(3) The superscripts, a, b, and c, denote significance at 0.10, 0.05, and 0.01 levels, respectively.

## Chapter 3

# An Empirical Structural Model of Productivity and Decision on the Conservation Reserve Program

### 1. Introduction

Agricultural production comprises about half of the total land use in the US (Lubowski et al. 2006). Agriculture is one of the areas most subject to the government interventions, e.g., from enhancing farmers' incomes by commodity price support to improving national food insecurity. Studying farmers' land use decisions is of interest because it often provides public implications that are fundamental to agricultural policy. The impacts of government policies are often reflected back into farmers' land use decisions. For example, the Conservation Reserve Program (CRP) provide incentives for farmers to retire their land from agricultural production, but at the same time agricultural subsidies increase the profitability of agricultural production leading to more intensive land use.

As one of the largest land conservation programs in US history, CRP was established by the Food Security Act of 1985 and reauthorized in every subsequent Farm Bill. In 2011, CRP paid \$1.7 billion to idle 31 million acres, which is roughly equivalent to the size of total farmland in Iowa (Kirwan et al. 2005). Under ten- to fifteen-year contracts, participants voluntarily retire environmentally sensitive farmland from agricultural production and receive annual rental payments from the U.S. Department of Agriculture. During the contract periods, farmland is converted to provide environmental benefits, for example, planting trees, wildlife cover, or other conservation uses. CRP is

also important to support landowners by providing a steady source of income (Sullivan et al. 2004). The environmental benefits provided by CRP are estimated to exceed the program costs (Ribaud et al. 1990; Feather et al. 1999).

In the last few years, however, CRP has faced a number of changes. As record-high crop prices significantly reduced landowners' incentives in the program, a large portion of CRP land has been put back into agricultural production. Figure 1 shows that CRP land went down by 10 million acres between 2007 and 2012, which accounted for 27% of total enrolled acres in 2007. Following the changes in market conditions, the 2014 Farm Bill allowed participants to terminate the contract prior to its expiration if the land has been enrolled more than five years and meets certain requirements.<sup>26</sup> The 2014 Farm Bill also reduced the cap of CRP acres to 24 million acres from 32 million acres allowed under the earlier Farm Bill.

Such changes raised an important policy question—to what extent do agricultural market conditions and CRP rental rates affect landowners' CRP incentives? Despite that there is an extensive literature on CRP, we have little evidence on how landowners react to changes in market prices and CRP rental rates in making CRP decisions. On one hand, CRP rental rates are considered as not high enough to keep up with the market rental rates under the recent strong crop prices. On the other hand, CRP contracts are viewed as long-term investments and steady income sources that should be less responsive to short-term market fluctuations. As far as crop prices stay high, finding an appropriate level of CRP rental rates that encourages the enrollments subject to the program costs will continue to be an important and challenging issue.

---

<sup>26</sup> See Stubbs (2014) for more details on the additional requirements.

We contribute to the literature in several fronts. First, we incorporate landowner's CRP decision in a model of land uses to investigate how farm-level CRP acreage is affected by agricultural prices and CRP payment. A novel identification strategy is developed to control for the endogeneity of CRP payment and self-selection of landowners into the program. These are corrected using the linkage between the landowner's bid on CRP payment and unobserved productivity in agricultural production. The intuition is that as the opportunity cost of taking land out of production increases with unobserved productivity, more productive farmers enrolled in CRP should be compensated by a higher rental payment. Our model takes explicit account of the endogeneity and self-selection issues. To the authors' knowledge, this paper is the first that models CRP decision in this way. Second, the structural model allows us to back out the time-varying farm-level productivity. We analyze the evolution of the farm productivity and further quantify the CRP payment premium to compensate more productive landowners. Third, we use the parameter estimates to simulate the land use changes at alternative agricultural prices and CRP payment across the Corn Belt states. A unique dataset is constructed from the USDA Census of Agriculture over 1997-2007.

The results show that a 10% increase in farm productivity leads to a 5.5% to 9.0% increase in CRP payment, or \$4.90 to \$8.02 higher CRP payment premium. If crop prices remain high and landowners have the flexibility to withdraw the contracts as allowed in the 2014 Farm Bill, the impact of crop prices on CRP enrollments would be significant. Although to some extent increasing CRP payment can offset the impact of high crop prices, it will require significant increases in the program costs. Moreover, benefits

provided by CRP land will change, as agricultural prices and CRP payment will have differentiated effects across the states.

The remainder of the paper is organized as follows. In the next section, we present the background of CRP followed by the literature review in Section 3. Sections 4 and 5 lay out the theoretical model and estimation method, respectively. In Section 6, we describe the data and Section 7 discusses the empirical results. In Section 8, we use the parameter estimates to simulate the land uses at alternative agricultural prices and CRP payment. Section 9 concludes.

## **2. Background**

Landowners participate in CRP through general or continuous signup. General signup involves a competitive bidding process for land to be enrolled. During each signup period, landowners submit bids by specifying a desired number of acres and a per acre rental payment. Landowners also specify a set of conservation practices, such as tree planting. Bids are ranked according to the Environmental Benefits Index (EBI), a scoring system that balances the environmental benefits against the costs. As currently structured, the benefit score is equal to the sum of five factors, including wildlife benefits, water quality, on-farm benefits from reduced soil erosion, air quality, and enduring benefits beyond the contract period. The cost score penalizes bids that propose higher rental rates or elect to share the costs of conservation practices with the government. Total EBI score is calculated by summing the benefit and cost scores.

Landowners can influence their EBI scores through proposed conservation practices or rental rates. For example, the EBI scores increase if landowners choose more valuable conservation practices or bid below the maximum rental rates. Also, if

landowners elect to share the costs of establishing conservation practices, the government shares up to the half of the costs. However, if they forgo the cost sharing, they receive extra points on their EBI scores. EBI threshold varies by signup and bids offered, therefore landowners know neither their EBI scores nor EBI threshold until bids are submitted. Landowners with the highest EBI scores are selected into the program until the target enrollment acreage is achieved. The most recent signup was held in 2013 and 88% of bids were accepted into the program with a national average rental payment of \$64/acre.<sup>27</sup>

The government sets the parcel-specific maximum rental rates (MRR) based on the county average market rental rate and the parcel-specific soil productivity. Soil rental rates are calculated by multiplying these two factors. The parcel-specific MRR is the weighted average of the soil rental rates for the proposed parcels.

While the majority of CRP land is enrolled under general signups, an increasing portion is enrolled under continuous signups.<sup>28</sup> In contrast to general signups, land can be enrolled under continuous signups at any time and is not subject to a competitive bidding process. Continuous signups are designed to target high priority conservation issues and offer additional financial incentives for qualified land to enroll. There are a number of initiatives under continuous signups, for example, Conservation Reserve Enhancement Program (CREP) and Farmable Wetland Program (FWP). In 2014, the average CRP

---

<sup>27</sup> The lowest acceptance rate was 71% in Wyoming with average rental payment of \$30/acre. The lowest average rental payment was \$27/acre in Pennsylvania. Rental payment varied significantly across the states, with the highest of \$194/acre in Iowa.

<sup>28</sup> In 2014 77% of CRP land was enrolled under general signups and the rest was enrolled under continuous signups (Stubbs 2014).

rental rates were: \$51/acre for the general signup, \$102, \$140 and \$114/acre for the non-CREP, CREP, and FWP continuous signup, respectively.

CRP contracts vary in duration, but most are for 10 to 15 years under both general and continuous signups. Landowners either apply to re-enroll in the program or let the contract expire at the end of the contracts. Opt-out was limited before 2014. Participants wishing to terminate the contracts early had to repay all the funds they received with interest and an additional penalty of 25% of rental payment. The 2014 Farm Bill, however, allowed participants to terminate the contracts prior to the expiration without penalty, if land was in CRP longer than five years and meets additional requirements.<sup>29</sup>

### **3. Literature Review**

The performance of CRP to achieve multiple environmental objectives in a cost-effective manner has been studied extensively in the literature (Jacobs et al. 2014). For example, Fleming (2004) finds that an increase in CRP acreage is associated with an improved quality of environment, suggesting a positive impact of CRP on the environment.

Jacobson (2014) finds negative temporal spillovers on the extensive margin, meaning that as CRP improves the quality of marginal land, it causes more land to be farmed in the future that potentially offset the environmental benefits provided by CRP land. However, temporal spillovers on the intensive margin are found to be positive. Farmed land previously in CRP was more likely to use environmental friendly practices. Kirwan et al. (2005) examined the cost-effectiveness of the CRP bidding process. Using the CRP signup data from 1997 to 2003, the authors find that CRP participants received rental payments higher than their reservation rents, and the associated premiums have increased

---

<sup>29</sup> For more details on the additional requirements, see Stubbs (2014).

over time accounting for 10%-40% of total rental payments. The authors argue that cost savings generated by adopting more efficient mechanisms could be substantial. Claassen et al. (2008) conclude that the effectiveness of CRP may have not been fully realized and the program failed to achieve the goal in a cost effective manner.

A number of studies investigate factors affecting landowners' CRP enrollment decisions. Chang and Boisvert (2009) distinguish the whole-farm and partial-farm participants of CRP using data from the 2001 Agricultural Resource Management Survey (ARMS). Whole-farms are those who enroll acres in CRP but produce no agricultural outputs. Partial-farms are with acres in both CRP and agricultural production. Partial-farm participants were more likely to accept greater variability of income associated with agricultural production. While CRP acreage enrolled by whole-farms is negatively associated with CRP payment, the relationship between CRP acreage and CRP payment is positive for partial-farms in areas with high soil quality and negative in areas with low soil quality. This is in contrast to findings in Fleming (2004) and Goodwin et al. (2004), where CRP payment is found to negatively affect CRP acreage enrollment. Hellerstein and Malcolm (2011) examine impacts of several factors on CRP, including acreage to be enrolled, EBI scores, and rental payment. The results highlight the sensitivity of CRP enrollment to crop prices. The results indicate that high crop prices in the absence of increases in rental rates would induce one-third of CRP land to be put back into agricultural production. Although the target acreage could be met by moderately increasing rental rates, it would be achieved at lower EBI scores because productive but sensitive land would tend to leave the program first.

#### 4. Model

Built on the model of land uses in Fezzi and Bateman (2011), we model landowner's CRP acreage choice under the framework of profit-maximizing land use decisions. Let  $\mathbf{p}$  and  $\mathbf{r}$  denote the respective output and input prices,  $L$  the total land acres in farm, and  $\omega$  the farm-specific productivity associated with agricultural production, which is known to farmers but not to the econometrician. The total land  $L$  is allocated to  $h$  uses. The profit function associated with the optimal land allocations is written as:

$$(1) \quad \Pi(\mathbf{p}, \mathbf{r}, \omega, L) = \underset{l_1, \dots, l_h}{\text{Max}} \left\{ \Pi(\mathbf{p}, \mathbf{r}, \omega, l_1, \dots, l_h) \mid \sum_{i=1}^h l_i = L \right\}.$$

The profit-maximizing input demand and output supply are derived using the Hotelling's lemma:

$$(2) \quad -X_i(\mathbf{p}, \mathbf{r}, \omega, L) = \frac{\partial \Pi(\mathbf{p}, \mathbf{r}, \omega, L)}{\partial r_i},$$

$$(3) \quad Q_i(\mathbf{p}, \mathbf{r}, \omega, L) = \frac{\partial \Pi(\mathbf{p}, \mathbf{r}, \omega, L)}{\partial p_i}.$$

Similarly, the optimal acreage allocations are obtained by equalizing the marginal rents or shadow prices of each land uses.

$$(4) \quad \frac{\partial \Pi(\mathbf{p}, \mathbf{r}, \omega, \bar{l}_1, \dots, \bar{l}_h)}{\partial l_i} = \frac{\partial \Pi(\mathbf{p}, \mathbf{r}, \omega, \bar{l}_1, \dots, \bar{l}_h)}{\partial l_j}.$$

For empirical estimation, we specify eqn. (1) using the flexible Normalized Quadratic (NQ) profit function, which has the advantage of incorporating negative profits that are typically not allowed in other flexible functional forms, such as the translog (Fezzi and Bateman 2011).

Denote  $\mathbf{x} = (p_1, \dots, p_m, r_1, \dots, r_n)$  a vector of output and input prices and  $\mathbf{s} = (l_1, \dots, l_{h-1}, L)$  a vector of  $h-1$  land allocations and total land acres, respectively. The last land use  $l_h$  is dropped out of the profit function as it can be calculated as  $L - \sum_{i=1}^{h-1} l_i$ .

The NQ profit function is written as

$$(5) \quad \begin{aligned} \Pi = & \alpha_0 + \sum_{i=1}^{m+n} \alpha_i x_i + \frac{1}{2} \sum_{i=1}^{m+n} \sum_{j=1}^{m+n} \alpha_{ij} x_i x_j + \sum_{i=1}^h \beta_i s_i + \frac{1}{2} \sum_{i=1}^h \sum_{j=1}^h \beta_{ij} s_i s_j \\ & + \sum_{i=1}^{m+n} \sum_{j=1}^h \gamma_{ij} x_i s_j + \delta_0 \omega + \sum_{i=1}^{m+n} \delta_i x_i \omega + \sum_{i=1}^h \varphi_i s_i \omega \end{aligned}$$

From eqn. (5), the optimal output supply can be written as

$$(6) \quad Q_i = \frac{\partial \Pi}{\partial x_i} = \alpha_i + \sum_{j=1}^{m+n} \alpha_{ij} x_j + \sum_{j=1}^h \gamma_{ij} s_j + \delta_i \omega.$$

The optimal land use allocations are determined by solving the following equations:

$$(7) \quad \frac{\partial \Pi}{\partial s_i} = \frac{\partial \Pi}{\partial s_j} \Leftrightarrow \beta_i + \sum_{j=1}^h \beta_{ij} s_j + \sum_{j=1}^{m+n} \gamma_{ij} x_j + \varphi_i \omega = \beta_j + \sum_{i=1}^h \beta_{ij} s_i + \sum_{i=1}^{m+n} \gamma_{ij} x_i + \varphi_j \omega \quad (i, j = 1, \dots, h-1).$$

Eqn. (7) together with the constraint that the sum of land uses is equal to total land acres provide the unique solution for  $h$  land allocations. Eqns. (6) and (7) specify a model with embedded structural parameters from which optimal output supply and land use allocations are decided upon agricultural prices, CRP payment, total land, and agricultural productivity. These two equations are, however, not sufficient to recover all the structural parameters of the profit function specified in eqn. (5). Without the profit data, in the subsequent analysis we focus on investigating landowners' CRP decisions among competing land uses by jointly estimating eqns. (6) and (7).

## 5. Estimation

### 5.1. The Output Supply Function

We consider three land use alternatives: (i) crop production  $l^a$ , (ii) CRP  $l^c$ , and (iii) remaining other land uses  $l^m$  (e.g., pasture, woodland, and livestock facilities). For simplification, all land for crop production is combined into  $l^a$ . Agricultural output and input prices are also aggregated into the composite output and input prices. Now the vector  $\mathbf{x}$  includes the composite agricultural output price  $p$ , input price  $r$ , and CRP payment per acre  $b$ , i.e.,  $\mathbf{x} \equiv (p, r, b)$ . For farmer  $k$  in state  $s$  at time  $t$ , the output supply function is specified as:

$$(8) \quad Q_{kt} = \alpha_1 + \alpha_{11}p_{st} + \alpha_{12}r_{st} + \alpha_{13}b_{kt} + \gamma_1L_{kt} + \delta_1\omega_{kt} + v_{kt}.$$

where  $v_{kt}$  is the i.i.d. error term. Assuming that farmers are price takers in agricultural output and input markets, the state-level prices  $p_{st}$  and  $r_{st}$  are used. In eqn. (8), the CRP payment  $b_{kt}$  is endogenous as it is potentially correlated with the unobserved productivity  $\omega_{kt}$ . That is, farmers with higher productivity should bid higher CRP payment to compensate higher opportunity costs of CRP participation, i.e., foregone profit from agricultural production. To deal with the endogeneity, we specify the CRP payment function as:

$$(9) \quad b_{kt} = b(\omega_{kt}, m_{ct}, L_{kt}, \mathbf{z}_{kt}).$$

where  $m_{ct}$  denotes the county-level market rental rate and  $\mathbf{z}_{kt}$  the farm-level demographic variables, including, e.g., operator's age and days worked off farm. In eqn.(9), the CRP payment is decided upon the county market rental rate and farm productivity. It is consistent with the fact that the parcel-specific maximum rental rate (MRR) is set based

on the county market rental rate and the soil productivity of proposed parcels. Eqn. (9) also includes total farmland, and a set of exogenous variables potentially affecting CRP payment in  $\mathbf{z}_{kt}$ . Eqn. (9) can be a general function but is required to be invertible, i.e., the monotonicity of CRP payment in productivity holds. Then we rely on the inverse of eqn. (9), i.e.,  $\omega(b_{kt}, m_{ct}, L_{kt}, \mathbf{z}_{kt})$ , to index the unobserved productivity. Plugging  $\omega(b_{kt}, m_{ct}, L_{kt}, \mathbf{z}_{kt})$  into eqn. (8), we rewrite the output supply function as:

$$(10) \quad Q_{kt} = \alpha_1 + \alpha_{11} p_{st} + \alpha_{12} r_{st} + \alpha_{13} b_{kt} + \gamma_1 L_{kt} + \delta_1 \omega(b_{kt}, m_{ct}, L_{kt}, \mathbf{z}_{kt}) + v_{kt}.$$

Note that the coefficients  $\alpha_{13}$  and  $\gamma_1$  are not separated from the coefficients of  $b_{kt}$  and  $L_{kt}$  in  $\omega(b_{kt}, m_{ct}, L_{kt}, \mathbf{z}_{kt})$  and therefore not identifiable in eqn. (10). Another complication is that eqn. (10) does not take into account that farmers' self-selection into CRP, and  $b_{kt}$  is only revealed for the CRP participants. We deal with this attrition issue by defining an indicator function:

$$(11) \quad B_{kt} = \begin{cases} 1 & \text{if } B_{kt}^* \geq \underline{B}_{kt}^*(m_{ct}, L_{kt}, \mathbf{z}_{kt}, \mathbf{c}_{ct}) \\ 0 & \text{otherwise} \end{cases}$$

where  $B_{kt}$  is equal to one if farmer  $k$  participates in the program at time  $t$ , and zero otherwise. In eqn. (11), the decision to participate in CRP is affected by the county market rental rate, total farmland, farm demographic variables in  $\mathbf{z}_{kt}$ , and the county agro-climatic conditions in  $\mathbf{c}_{ct}$ .

Finally, the coefficients  $\alpha_{13}$  and  $\gamma_1$  are identified by placing a structural assumption on the evolution of the productivity, that the productivity follows a first-order Markov process:

$$(12) \quad \omega_{kt} = E[\omega_{kt} | \omega_{kt-1}, B_{kt} = 1] + \varepsilon_{kt} = g(b_{kt-1}, m_{ct-1}, L_{kt-1}, \mathbf{z}_{kt-1}, \lambda_{kt}) + \varepsilon_{kt}.$$

Here  $\lambda_{kt}$  is the inverse mills ratio (IMR) obtained from the Probit regression of eqn. (11), and  $\varepsilon_{kt}$  represents the productivity innovations in the current period. Plugging eqn. (12) into eqn. (8), we have

$$(13) \quad Q_{kt} = \alpha_1 + \alpha_{11}p_{st} + \alpha_{12}r_{st} + \alpha_{13}b_{kt} + \gamma_1 L_{kt} + \delta_1 g(b_{kt-1}, m_{ct-1}, L_{kt-1}, \mathbf{z}_{kt-1}, \lambda_{kt}) + u_{kt}$$

where  $u_{kt} \equiv v_{kt} + \varepsilon_{kt}$ . Following Wooldridge (2009), eqn. (10) and (13) are jointly estimated by the Generalized Method of Moment (GMM) using the different orthogonality conditions across the equations:

$$(14) \quad E(v_{kt} | p_{st}, r_{st}, b_{kt}, L_{kt}, \Theta) = 0$$

$$(15) \quad E(u_{kt} | p_{st}, r_{st}, b_{kt-1}, L_{kt}, \Psi) = 0$$

where vectors  $\Theta$  and  $\Psi$  contain the instrumental variables for  $\omega(b_{kt}, m_{ct}, L_{kt}, \mathbf{z}_{kt})$  and  $g(b_{kt-1}, m_{ct-1}, L_{kt-1}, \mathbf{z}_{kt-1}, \lambda_{kt})$ , respectively. The error term  $v_{kt}$  in eqn. (10) is i.i.d. by construction, so the variables act as their own instruments. Although eqn. (10) does not identify  $\alpha_{13}$  and  $\gamma_1$ , the coefficients are identified in eqn.(13). In eqn. (13), because that  $b_{kt}$  is potentially correlated with the productivity innovations  $\varepsilon_{kt}$  in  $u_{kt}$ , we use  $b_{kt-1}$  as the instrument. By the construction of the first-order Markov in eqn. (12),  $b_{kt-1}$  is not correlated with  $\varepsilon_{kt}$ .

## 5.2. The Acreage Equations

Solving eqn. (7) for land used in agricultural production  $l_{kt}^a$  and CRP  $l_{kt}^c$ , we have

$$(16) \quad l_{kt}^a = \theta_1 + \theta_{11}p_{st} + \theta_{12}r_{st} + \theta_{13}b_{kt} + \theta_{14}L_{kt} + \theta_{15}\omega_{kt} + v_{kt}^a$$

$$(17) \quad l_{kt}^c = \theta_2 + \theta_{21}p_{st} + \theta_{22}r_{st} + \theta_{23}b_{kt} + \theta_{24}L_{kt} + \theta_{25}\omega_{kt} + v_{kt}^c$$

The error terms  $v_{kt}^a$  and  $v_{kt}^h$  account for the measurement errors. The estimation method is directly analogous to that for the output supply function discussed above. Plugging the inverse of eqns. (9) and the dynamics of the productivity in eqn. (12) into eqns. (16) and (17), we have the following four equations for the two land uses.

$$(18) \quad l_{kt}^a = \theta_1 + \theta_{11}p_{st} + \theta_{12}r_{st} + \theta_{13}b_{kt} + \theta_{14}L_{kt} + \theta_{15}\omega(b_{kt}, L_{kt}, m_{ct}, \mathbf{z}_{kt}) + v_{kt}^a,$$

$$(19) \quad l_{kt}^a = \theta_1 + \theta_{11}p_{st} + \theta_{12}r_{st} + \theta_{13}b_{kt} + \theta_{14}L_{kt} + \theta_{15}g(b_{kt-1}, L_{kt-1}, m_{ct-1}, \mathbf{z}_{kt-1}, \lambda_{kt}) + u_{kt}^a,$$

$$(20) \quad l_{kt}^c = \theta_2 + \theta_{21}p_{st} + \theta_{22}r_{st} + \theta_{23}b_{kt} + \theta_{24}L_{kt} + \theta_{25}\omega(b_{kt}, L_{kt}, m_{ct}, \mathbf{z}_{kt}) + v_{kt}^c,$$

$$(21) \quad l_{kt}^c = \theta_2 + \theta_{21}p_{st} + \theta_{22}r_{st} + \theta_{23}b_{kt} + \theta_{24}L_{kt} + \theta_{25}g(b_{kt-1}, L_{kt-1}, m_{ct-1}, \mathbf{z}_{kt-1}, \lambda_{kt}) + u_{kt}^c.$$

Eqns. (18)-(21) are jointly estimated with eqns. (10) and (13) relying on the moment conditions in eqns. (14) and (15).

## 6. Data

We construct a unique panel dataset of farms in major Corn Belt states using the confidential USDA Census of Agriculture over 1997-2007. The census is the only farm-level panel data available for the US agriculture. The census is conducted every five years, so the constructed dataset consists of farms in the year of 1997, 2002, and 2007. The census surveys all the farms that produce or normally would have produced \$1,000 or more of agricultural products during the census year. Farms are mandated to participate in the census, and are linked longitudinally using a numeric identifier. In this paper, the Corn Belt states include Kansas, Nebraska, Missouri, Iowa, Illinois, Indiana, Wisconsin, Michigan, Minnesota, and Ohio. These states cultivate corn and soybean as the predominant crops and account for a significant portion of agricultural production in the US. We restrict our sample to crop farms, as the primary purpose of CRP is to retire

farmland from crop production.<sup>30</sup> Crop farms are defined as farms that generate at least 80% of total revenue from crop production.

The data on farmland uses ( $l_{kt}^a$ ,  $l_{kt}^c$ , and  $L_{kt}$ ), CRP rental payment ( $b_{kt}$ ), operator's age and days worked off farm ( $z_{kt}$ ) are obtained from the census. Since the census does not distinguish CRP land from land enrolled in the Wetland Reserve Program (WRP),<sup>31</sup> we use the county-level CRP and WRP data and exclude farms in the counties where more than 10% of CRP or WRP land is enrolled in WRP. The data for the other variables are obtained from publically available sources. Specifically, state-level corn and soybean prices are obtained from the USDA Agricultural Prices Summary<sup>32</sup>. The composite output price ( $p_{st}$ ) is calculated as the weighted average of corn and soybean prices with the weights being the shares of harvested acres for individual crops at the state level. Regional-level nitrogen prices ( $r_{st}$ ) are also available from the USDA Agricultural Prices Summary. The county-level market rental rates ( $m_{ct}$ ) are collected from the county-level Census of Agriculture. Finally, the county-level agro-climatic variables included in  $c_{ct}$  are the growing degree-days (GDDs), which are constructed by Schlenker and Roberts (2009), and the agricultural capability soil index calculated from the 1997 National Resources Inventory (NRI).

---

<sup>30</sup> In 2004, pasture land accounted for less than 1% of the total CRP enrollment.

<sup>31</sup> CRP and WRP are the two primary agricultural land retirement programs in the US, although WRP is smaller in scale with its annual expenditures being roughly one tenth of the CRP's.

<sup>32</sup> The USDA Agricultural Prices Summary is available on <http://usda.mannlib.cornell.edu/MannUsda/viewDocumentInfo.do?documentID=1003>.

Following the definition for no-CRP, partial, and whole farms in Chang and Boisvert (2009) that we described in the literature review, Table 1 provides descriptive statistics for the three groups of farms. The last column reports the Wald Chi-square statistics to test the null hypothesis that the sample means of the variable in each row are equal across the groups. Table 1 shows that no-CRP farms are generally located in areas with relatively higher output, fertilizer and land rental prices, the differences of which are statistically significant across the groups. Partial farms typically have more acres of farmland and cultivate more cropland than no-CRP farms. But we fail to reject that no-CRP farms produce the same amount of crop as partial farms, which implies that crop yield is higher for no-CRP farms than partial farms.

On average, operators in whole farms are the oldest among the groups followed by partial and no-CRP farms. Days worked off-farm is measured as a categorical variable ranging from 1 to 5: 1 for 0, 2 for 1-49, 3 for 50-99, 4 for 100-199 and 5 for 200 or more days. Days worked off farm are the highest for whole farms and the lowest for partial farms. Both the growing degree-days (GDDs) and soil index are greater in whole farms than others, suggesting that agro-climatic conditions are the most beneficial in the areas where the whole farms are located. Compared with partial farms, whole farms enroll more acres in CRP but receive lower payment per acre. However, CRP payment is higher than the market rental rate for both partial and whole farms.

Table 2 summarizes the characteristics of partial farms over the sample period. Between 2002 and 2007, output and fertilizer prices were almost doubled and the market rental rates increased by 50%. Following the market trends, both total land and cultivated cropland increased by 50%, while CRP land decreased by 10%. Although increased

significantly between 2002 and 2007, the market rental rates are lower than the CRP payment during the whole sample period. We observe both crop production and yield increased, but days worked off farm were lowered over 2002-2007. Higher GDDs and lower average soil quality were observed in regions where partial farms were in 2007, which implies that partial farms in relatively high soil quality regions terminated or did not renew CRP contracts, or more partial farms in low soil quality regions entered into the program when commodity crop prices were record-high in 2007.

## **7. Empirical Results**

In this section, we start with the estimation results for landowners' binary CRP decision followed by output supply and land use equations. Then, using the farm-level productivity estimates, we measure the CRP payment premium compensated to high productivity landowners.

### **7.1. Binary CRP Participation Decision**

The indicator function specified in eqn. (11) is estimated to investigate the landowners' binary decisions on CRP participation. Table 3 presents the marginal effects and the elasticities of the Probit regression. The results indicate that the county market rental rate and the growing-degree days (GDDs) negatively affect CRP participation, while total farmland, operator's age, days worked off farm, and soil index are positively associated with the predicted probability.

Specifically, the predicted probability of participation decreases by 0.19% if the county market rental rate increases by 1%, or \$0.76 at the sample means. This makes sense as the market rental rate reflects the profitability of agricultural production or the opportunity costs of CRP participation. The predicted probability increases by 0.05% if

farm size represented by total farmland increases by 1%, or 6 acres evaluated at the sample means. Also, the probability increases by 0.46% if the operator's age increases by 0.5 years, which implies that the CRP enrollment decision is related with landowner's life cycle in that putting land in CRP is a way to prepare landowners for retirement.

Although both soil quality and GDDs are indicators of favorable regional agro-climatic conditions, they affect CRP participation in different ways. While a 1% increase in soil quality is associated with a 0.55% increase in the predicted probability, a 1% increase in GDDs decreases the probability by 0.36%. The results also indicate that days worked off farm increase the predicted probability by 1-2%. This suggests that some landowners may participate in CRP for reducing the need for farm labor.

## **7.2. Optimal Output Supply and Land Allocations**

Table 4 presents the joint estimation results for the output supply and land use equations. All coefficients are intuitively signed and statistically significant. Output supply and cropland acres increase with output price, and decrease with fertilizer price and CRP payment. On the other hand, CRP acreage is positively associated with fertilizer price and CRP payment, and negatively affected by output price. Total farmland acres show positive effects in all the equations. Although existing studies in the literature find negative relationship between CRP payment and enrollment (see, e.g., Fleming 2004; Chang and Boisvert 2009), we confirm that after controlling for confounding factors for CRP payment, the net effect of CRP payment is positive in CRP enrollment and is negative in cropland and output supply. In a latter section, we use the estimated coefficients in Table 4 to simulate the land uses at alternative input and output prices as well as CRP payment.

### 7.3. Payment Premiums of High-Productivity Farms

In the estimation of output supply and land use equations of (8), (16), and (17), the unobserved productivity is partially identified multiplied by the associated coefficients, i.e.,  $\delta_1 \omega_{kt}$ ,  $\theta_{15} \omega_{kt}$ , and  $\theta_{25} \omega_{kt}$ . Based on the estimates of the output supply function, we calculate the relative productivity as<sup>33</sup>

$$(22) \quad \omega_{kt}^R = \frac{\delta_1 \omega_{kt}}{\frac{1}{K} \frac{1}{T} \sum_k \sum_t \delta_1 \omega_{kt}} = \frac{\omega_{kt}}{\bar{\omega}}.$$

Here the relative productivity is equal to 1 for average farms, greater and lower than 1 for higher- and lower-productivity farms than average, respectively. Figure 2 shows the distribution of relative productivity for each sample period. Most of farms are clustered around average, while a small number of farms exhibit very high productivity. Although the productivity generally increases over time, it became more dispersed when farms with substantially higher productivity emerged in 2007.

Given the CRP payment function specified in eqn. (9), we regress CRP payment on the relative productivity, the county market rental rate, total farmland, and farmer's demographic variables as the following:

$$(23) \quad b_{kt} = \tau_1 + \tau_{11} \omega_{kt}^R + \tau_{12} m_{ct} + \tau_{13} L_{kt} + \mathbf{z}_{kt} \boldsymbol{\sigma} + v_{kt}.$$

In eqn. (23), all variables are log transformed, so the corresponding coefficients are interpreted as elasticities.

Table 5 presents the regression results where the standard errors are clustered at farm-level to control for potential heteroskedasticity and serial correlation. Column I

---

<sup>33</sup> The productivity estimates for a total of 119 observations were negative and thus dropped in the subsequent analysis.

includes only the relative productivity as the explanatory variable. The elasticity of 0.55 indicates a 5.5% increase in CRP payment responding to a 10% increase in relative productivity. Evaluated at the sample means this corresponds to \$4.9 higher CRP payment. When all other variables are included in column II, we find a 10% increase in productivity leads to a 7.6% premium of the CRP payment. For robustness check, in columns III-VIII we add in various sets of region and time fixed effects and find similar results of CRP payment premium induced by a 10% increase in productivity, i.e., 7.5%-9.0%.<sup>34</sup>

We also find significant and robust impacts of the other variables. Specifically, CRP payment increases with the county market rental rate, operator's age, and days worked off farm, and decreases with total farmland. The impact of a 10% increase in the county market rental rate, operator's age, and days worked off farm on the payment premium varies from 0.6%-4.2%, 8.1-13%, and 0.6-2.3%, respectively. A 10% increase in total farmland reduces CRP payment by 1.7%-3.4%.

## **8. Policy Simulation**

In this section, we quantify to what extent agricultural prices and CRP payment influence the program enrollment and costs using the estimated coefficients in Table 4 and data in 2007 and 2012. As shown in Figure 1, increasing agricultural commodity prices in 2007-2012 significantly reduced land enrolled in CRP. More importantly, the impacts were differentiated across the states as presented in Figure 3. Although CRP acreage was dropped in all Corn Belt states, it was reduced the most in Kansas, followed by Nebraska

---

<sup>34</sup> Given the lack of within-farm variations, the regressions with farm-level fixed effects have limited power to identify the payment premium.

where a large amount of land was enrolled at relatively low rental rates. The acreage reductions were also significant in Iowa and Missouri where both CRP acreage and rental rates are relatively higher than other states. Relatively insignificant changes occurred in Indiana and Ohio. Changes in the program costs were also differentiated across the states as presented in Figure 4. Between 2007 and 2012, the program costs increased in Iowa, Illinois, Missouri, and Ohio as increases in rental rates offset the reduced acres. Other states experienced the reduced program costs with the biggest reduction in Kansas.

We simulate the following seven scenarios: (i) a 10% increase in output price holding other prices at the 2007 level, (ii) a 10% increase in input price holding other prices at the 2007 level, (iii) a 10% increase in CRP payment holding other prices at the 2007 level, (iv) increasing output price to the level of 2012 holding other prices at the 2007 level, (v) increasing input price to the level of 2012 holding other prices at the 2007 level, (vi) increasing CRP payment to the level of 2012 holding other prices at the 2007 level, and (vii) increasing output and input prices and CRP payment to the levels of 2012. Scenarios (i)-(iii) provide the benchmark assessment for the impact of input/output prices and CRP payment and scenarios (iv)-(vi) decompose the aggregated changes over 2007-2012 to each factors. The purpose of scenario (vii) is to compare the model predictions with the actual changes for validation.

Table 6 summarizes the data used for the simulation. Over the period of 2007-2012, output price increased by 53% to 76% across the states. We apply the national average increase of 34% to all the states because regional nitrogen prices are not available in 2012. During the same period, the state average CRP payment increased by 3% to 24%, the lowest of which was observed in Kansas and the highest in Indiana and Iowa.

Using the coefficient estimates in Table 4, for each state we calculate the rate of counterfactual changes in CRP acreage between 2007 and 2012:

$$(24) \quad \Delta \equiv \frac{\sum_{k \in s} \hat{l}_{kt}^c - \sum_{k \in s} \hat{l}_{kt}^c}{\sum_{k \in s} \hat{l}_{kt}^c},$$

where  $\sum_{k \in s} \hat{l}_{kt}^c$  denotes the sum of predicted CRP acreage of farms in state  $s$  in 2007, and  $\sum_{k \in s} \hat{l}_{kt}^c$  represents the sum of predicted CRP acreage of farms in state  $s$  at alternative price levels. For each state, counterfactual acreage change is calculated by multiplying the rate of counterfactual change  $\Delta$  in eqn. (24) by actual CRP acreage in state  $s$  in 2007.

Table 7 shows the simulation results of the seven scenarios with the actual CRP acreage changes in the last column. Comparing the first three scenarios, columns (i) to (iii), we find that CRP payment has in general the largest impact on CRP acreage followed output and input price, except that the impact of output price is largest in Kansas and Nebraska. Across the states, CRP enrollment is affected the most in Iowa, where a 10% increase in output price decreases CRP acreage by 0.81 million acres and a 10% increase in input price and CRP payment increases CRP acreage by 0.61 and 1.33 million acres, respectively. The impact is smallest in Michigan, which have the least land acres in CRP.

Columns (iv) to (vii) show the counterfactual changes in CRP acreage at 2012 prices and indicate heterogeneous impacts across the states. For example, in column (iv), CRP acreage drops by 0.24-1.85 million acres due to high output prices of 2012, and in column (v), high input prices of 2012 pushes up CRP acreage by 0.23 to 2.06 million acres across the states. When CRP payment increases to the 2012 level holding output

and input prices at the baseline in 2007, in column (vi), the results show that CRP acreage increases the least in Kansas, by 0.05 million acres, and the most in Iowa by 2.98 million acres.

Column (vii) summarizes the simulation results where input/output prices and CRP payment all increase to the 2012 level. Compared with the actual changes presented in the last column, the changes predicted in our model are generally consistent with the observed changes. Small impacts on CRP acreage are predicted in areas where actual changes were relatively insignificant, including Indiana, Michigan, and Ohio, and large impacts are predicted in regions where significant changes in CRP acreage were actually observed, such as Kansas, Nebraska, and Iowa. Figure 5 maps the results for the scenario (vii) and shows essentially the same spatial patterns with the actual changes presented in Figure 3.

We note that the model tends to predict larger effects on CRP acreage than actually observed during the period of record-high crop prices. Our model overestimates the impacts of market prices and CRP payment on CRP acreage for two reasons. First, we do not consider the opt-out of CRP contracts, which was limited until the 2014 Farm Bill. Second, our sample only consists of partial farms with acres in both CRP and agricultural production. Partial farms would be more responsive to changes in market conditions and CRP payment than whole farms, which have all the acres enrolled in CRP and produce no farm products. In that sense the simulation results provide the upper bounds of impacts of the factors on CRP acreage. Given that the 2014 Farm Bill allows participants to opt-out without penalty under certain conditions, our results suggest that impacts on CRP acreage

will be significant if agricultural prices stay relatively high and landowners can flexibly withdraw CRP contracts.

In Table 8, we investigate two scenarios to predict additional CRP acreage and costs that would have been incurred if CRP payment were higher by 10% and 20% than the 2012 level holding output and input prices at those in 2012. Figures 6 and 8 visualize the results for the two scenarios. Large increases in CRP acres are found in Minnesota, Iowa, Missouri, and Illinois, where relatively large acreage is enrolled in CRP at high rental payment. Nebraska, Kansas, and Wisconsin show moderate increases in CRP land. The results suggest that increasing CRP payments can partially offset impacts of high commodity crop prices in states where rental rates are high, but effectiveness of CRP payment varies across the states.

Moreover, as shown in Table 8, significant increases in the program costs will be needed to offset impacts of high crop prices. Figures 7 and 9 present the additional program costs. The cost increases vary across the states, with the most significant increases in Iowa, Illinois, Minnesota, and Missouri. Although acreage increases are expected to be small in Indiana, Ohio, and Michigan, the extra program costs are not negligible given relatively high rental rates in these states.

## **9. Conclusion**

To better understand driving factors of CRP enrollment, we incorporate Corn Belt landowners' CRP decisions in a model of land uses. The land use model carefully addresses two econometric issues, the endogeneity of CRP payment and self-selection of landowners, using the linkage between landowners' bids on CRP payment and the unobserved agricultural productivity. We hypothesize that higher-productivity

landowners should submit higher payment bids to offset higher opportunity costs of CRP participation, i.e., the foregone profits from agricultural production. Self-selection arises as CRP payment is revealed only for participants and is corrected by incorporating Heckman correction into the model.

In landowners' binary decisions on CRP participation, we find that total farmland acres, operator's age, days worked off farm, and soil quality all show positive effects. In contrast, the county market rental rate and growing degree-days are negatively associated with CRP participation. Farm-level productivity is calculated as residual from the estimation of land use model. The results confirm our hypothesis, where higher farm productivity is associated with a premium of CRP payment. Specifically, we find that a 10% increase in farm productivity leads to \$4.90 to \$8.02 higher CRP payment at sample means.

Furthermore, we simulate impacts of alternative agricultural prices and CRP payment on CRP acreage and costs across the major Corn Belt states. The simulation results indicate that impacts on CRP would be significant if agricultural prices stay at high and landowners have flexibility to opt-out the contracts. Although increasing CRP payments can partially offset effects of high agricultural commodity prices, significant increases in the program costs will be needed.

We conclude with several caveats. First, our structural model doesn't distinguish subprograms under CRP due to data limitations. Second, our model treats all the non-participants in the same way, although reasons for not participating may be different across the non-participants, for example, they either did not consider CRP at all or submitted bids that were rejected. Such information is not available in the census. Third,

we deal with the endogeneity of CRP payment by assuming that there is a monotonic relationship between the payment and the unobserved productivity. Although the results indicate the significant relationship between the payment and the productivity to capture effects of unobservables, it is possible that in some cases the invertibility condition fails, especially if landowners condition their bids on the environmental benefits reflected in their EBI scores and such benefits are not captured in the productivity. Finally, a potential extension of our model to a dynamic setting would enable us to better understand inter-temporal land use decisions of landowners, which is necessary to address long-term aspects of CRP contracts.

Table 3-1. Descriptive Statistics of No-CRP, Partial, and Whole Farms.

	No-CRP Farms	Partial Farms	Whole Farms	Wald Chi-square
Output price index (\$/bu)	4.89 (1.21)	4.65 (1.12)	4.74 (1.15)	914.56 (0.00)
Fertilizer price (\$/ton)	260.59 (91.06)	245.65 (82.04)	256.09 (84.32)	665.25 (0.00)
Land rental price (\$/acre)	83.64 (44.59)	70.78 (43.78)	58.92 (41.73)	2686.19 (0.00)
Total land in farm (acre)	576.51 (760.78)	717.15 (996.99)	124.69 (174.63)	18651.48 (0.00)
Cropland cultivated (acre)	500.48 (658.43)	510.01 (773.95)	-	3.98 (0.04)
CRP land (acre)	-	74.00 (143.94)	124.69 (174.63)	393.42 (0.00)
CRP payment (\$/acre)	-	87.51 (134.52)	68.71 (64.39)	297.27 (0.00)
Crop produced (bu)	32606.61 (52724.56)	32136.17 (54058.45)	178.80 (1974.96)	1.73 (0.18)
Total output produced	33239.57 (53428.63)	32762.24 (55041.83)	237.95 (2273.30)	1.73 (0.18)
Crop yield (bu/acre)	60.92 (47.97)	58.17 (28.88)	-	92.63 (0.00)
Operator's age	54.07 (13.03)	56.92 (12.76)	61.08 (14.58)	1724.81 (0.00)
Days worked off farm	2.79 (1.80)	2.61 (1.77)	2.95 (1.87)	319.26 (0.00)
Growing degree days	1997.87 (243.78)	2013.91 (269.05)	2025.68 (312.46)	106.61 (0.00)
Soil index	2.35 (0.33)	2.45 (0.35)	2.56 (0.34)	2514.39 (0.00)
# Observations	39,420	50,828	4,979	

## Notes:

- (1) Standard deviations and p-values are in the parentheses in the first three, and in the last column, respectively.
- (2) Output price index is the weighted average of corn and soybean prices (\$/bu) at the state-level with the weights being the shares of the harvested acres for individual crops.
- (3) Days worked off farm is a categorical variable ranging from 1 to 5: 1 for 0, 2 for 1-49, 3 for 50-99, 4 for 100-199 and 5 for 200 or more days.

Table 3-2. Descriptive Statistics of Partial Farms.

	1997	2002	2007	Wald Chi-square
Output price index	4.20 (0.36)	3.82 (0.24)	6.21 (0.36)	542358.82 (0.00)
Fertilizer price (\$/ton)	219.65 (3.24)	179.77 (0.62)	363.09 (2.53)	82900000 (0.00)
Land rental price (\$/acre)	44.02 (30.18)	64.01 (37.29)	96.10 (46.76)	11568.21 (0.00)
Total land in farm (acre)	851.32 (975.85)	597.04 (909.32)	831.41 (1110.47)	771.61 (0.00)
Cropland cultivated (acre)	566.52 (679.38)	398.56 (693.14)	653.41 (904.71)	1090.45 (0.00)
CRP land (acre)	97.38 (170.68)	73.12 (133.53)	62.51 (142.09)	271.59 (0.00)
CRP payment (\$/acre)	68.76 (32.42)	84.59 (183.74)	102.37 (50.16)	4162.81 (0.00)
Crop produced (bu)	32282.38 (43495.28)	23916.59 (46056.57)	44923.97 (66888.91)	1279.35 (0.00)
Total output produced	32796.37 (44217.24)	24521.31 (47147.57)	45645.10 (67907.35)	1245.37 (0.00)
Crop yield (bu/acre)	52.12 (24.95)	54.97 (26.15)	65.72 (32.42)	1465.56 (0.00)
Operator's age	53.68 (12.70)	56.88 (12.81)	58.76 (12.15)	954.26 (0.00)
Days worked off farm	2.55 (1.74)	2.68 (1.81)	2.52 (1.71)	96.47 (0.00)
Growing degree days	2055.01 (277.96)	1999.49 (277.28)	2013.90 (247.85)	265.15 (0.00)
Soil index	2.47 (0.34)	2.48 (0.35)	2.39 (0.35)	688.74 (0.00)
# Observations	8,971	25,542	16,315	

## Notes:

- (1) Standard deviations and p-values are in the parentheses in the first three, and in the last column, respectively.
- (2) Output price index is the weighted average of corn and soybean prices (\$/bu) at the state-level with the weights being the shares of the harvested acres for individual crops.
- (3) Days worked off farm is a categorical variable ranging from 1 to 5: 1 for 0, 2 for 1-49, 3 for 50-99, 4 for 100-199 and 5 for 200 or more days.

Table 3-3. Probit Regression Results.

	Marginal effect	Elasticity
Land rental price	-0.001 (0.00004) <sup>c</sup>	-0.19 (0.006) <sup>c</sup>
Total land in farm	0.00005 (0.000002) <sup>c</sup>	0.05 (0.002) <sup>c</sup>
Operator's age	0.004 (0.0001) <sup>c</sup>	0.46 (0.01) <sup>c</sup>
Days worked off farm		
2	0.02 (0.005) <sup>c</sup>	
3	0.01 (0.008) <sup>b</sup>	
4	0.01 (0.006) <sup>c</sup>	
5	0.01 (0.004) <sup>c</sup>	
Soil index	0.13 (0.005) <sup>c</sup>	0.55 (0.02) <sup>c</sup>
Growing degree days	-0.0001 (0.000007) <sup>c</sup>	-0.36 (0.02) <sup>c</sup>
# Observations	89,732	89,732

Note:

(1) Standard errors are in the parentheses.

(2) The superscripts, a, b, and c denote significance at 0.10, 0.05, and 0.01 levels, respectively.

Table 3-4. GMM Estimation Results.

	Output supply	Cropland	CRP land
Constant	-8844.48 (6154.18)	-26.94 (62.93)	70.61 (30.53) <sup>b</sup>
Output price	6791.86 (486.54) <sup>c</sup>	31.44 (5.20) <sup>c</sup>	-13.89 (2.49) <sup>c</sup>
Fertilizer price	-203.55 (16.30) <sup>c</sup>	-0.95 (0.17) <sup>c</sup>	0.42 (0.08) <sup>c</sup>
CRP payment	-497.63 (149.97) <sup>c</sup>	-7.01 (1.64) <sup>c</sup>	3.06 (0.83) <sup>c</sup>
Total land in farm	10.21 (0.88) <sup>c</sup>	0.72 (0.01) <sup>c</sup>	0.01 (0.004) <sup>c</sup>
# Observations	18,328	18,328	18,328

Note:

(1) Standard errors are in the parentheses.

(2) The superscripts, a, b, and c denote significance at 0.10, 0.05, and 0.01 levels, respectively.

Table 3-5. Productivity Regression Results.

	Dependent variable: CRP payment							
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)
Productivity estimates	0.55 (0.004) <sup>c</sup>	0.76 (0.008) <sup>c</sup>	0.79 (0.01) <sup>c</sup>	0.83 (0.01) <sup>c</sup>	0.90 (0.01) <sup>c</sup>	0.75 (0.008) <sup>c</sup>	0.77 (0.009) <sup>c</sup>	0.76 (0.009) <sup>c</sup>
Land rental price		0.06 (0.004) <sup>c</sup>	0.12 (0.01) <sup>c</sup>	0.42 (0.02) <sup>c</sup>	1.80 (0.01) <sup>c</sup>	0.10 (0.005) <sup>c</sup>	0.15 (0.005) <sup>c</sup>	0.16 (0.005) <sup>c</sup>
Total land in farm		-0.18 (0.002) <sup>c</sup>	-0.23 (0.005) <sup>c</sup>	-0.25 (0.004) <sup>c</sup>	-0.34 (0.004) <sup>c</sup>	-0.17 (0.003) <sup>c</sup>	-0.18 (0.003) <sup>c</sup>	-0.18 (0.003) <sup>c</sup>
Operator's age		1.30 (0.007) <sup>c</sup>	0.89 (0.02) <sup>c</sup>	0.81 (0.01) <sup>c</sup>	0.28 (0.008) <sup>c</sup>	1.27 (0.007) <sup>c</sup>	1.25 (0.008) <sup>c</sup>	1.22 (0.008) <sup>c</sup>
Days worked off farm		0.23 (0.003) <sup>c</sup>	0.16 (0.004) <sup>c</sup>	0.15 (0.003) <sup>c</sup>	0.06 (0.002) <sup>c</sup>	0.22 (0.003) <sup>c</sup>	0.22 (0.003) <sup>c</sup>	0.22 (0.003) <sup>c</sup>
Fixed effect								
County			X	X				
State						X	X	
Time				X			X	
County-time					X			
State-time								X
R-squared	0.41	0.99	0.99	0.99	0.99	0.99	0.99	0.99
#Observations	38,071	38,071	38,071	38,071	38,071	38,071	38,071	38,071

Note:

(1) Robust standard errors clustered at farm-level are in the parentheses.

(2) The superscripts, a, b, and c denote significance at 0.10, 0.05, and 0.01 levels, respectively.

Table 3-6. Prices Used in Simulation.

	Baseline (2007)			2012		
	Output	Fertilizer	CRP payment	Output	Fertilizer	CRP payment
Illinois	6.07	364	103.36	10.20	487.76	120.40
Indiana	6.67	364	92.35	10.99	487.76	114.47
Iowa	6.25	364	106.16	10.09	487.76	131.57
Kansas	5.90	356	39.25	10.06	477.04	40.61
Michigan	6.14	364	75.89	9.40	487.76	90.53
Minnesota	6.21	364	60.28	10.03	487.76	69.85
Missouri	7.10	364	66.97	12.51	487.76	79.24
Nebraska	5.49	356	57.02	9.09	477.04	62.39
Ohio	6.88	364	102.06	11.37	487.76	122.06
Wisconsin	5.38	364	70.36	8.76	487.76	83.37

## Notes:

- (1) Output price index is the weighted average of corn and soybean prices (\$/bu) at the state-level with the weights being the shares of the harvested acres for individual crops.
- (2) 2007 fertilizer price is the nitrogen price (\$/ton) at regional-level. 2012 fertilizer price is obtained by multiplying the national increase rate of 34% by the price in 2007.
- (3) CRP payment (\$/acre) is the average CRP rental rate at state-level, which is calculated by dividing total rental payment by total enrolled acres.

Table 3-7. Changes in CRP Acreage (in million acres).

	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	Actual changes
Illinois	-0.33	0.24	0.40	-0.99	0.83	0.66	-0.53	-0.06
Indiana	-0.11	0.08	0.13	-0.29	0.26	0.31	-0.11	-0.04
Iowa	-0.81	0.61	1.34	-1.85	2.06	2.98	-0.55	-0.33
Kansas	-0.30	0.22	0.17	-1.82	0.75	0.05	-1.14	-0.74
Michigan	-0.09	0.07	0.11	-0.24	0.23	0.22	-0.06	-0.05
Minnesota	-0.51	0.41	0.57	-1.56	1.36	0.90	-0.71	-0.27
Missouri	-0.36	0.28	0.37	-1.39	0.89	0.65	-0.78	-0.31
Nebraska	-0.21	0.17	0.20	-0.89	0.55	0.19	-0.49	-0.35
Ohio	-0.12	0.10	0.17	-0.33	0.32	0.32	-0.12	-0.03
Wisconsin	-0.21	0.17	0.23	-0.56	0.53	0.41	-0.24	-0.24

## Notes:

(i) Output price increases by 10% holding other prices at the 2007 level.

(ii) Input price increases by 10% holding other prices at the 2007 level.

(iii) CRP payment increases by 10% holding other prices at the 2007 level.

(iv) Output price increases to 2012 price holding other prices at the 2007 level.

(v) Input price increases to 2012 price holding other prices at the 2007 level.

(vi) CRP payment increases to the payment in 2012 holding other prices at the 2007 level.

(vii) All variables increase to the 2012 prices.

Table 3-8. Additional CRP Acreage and Costs by Increased Payment.  
(in million acres; \$ millions).

	10%		20%	
	Acreage	Cost	Acreage	Cost
Illinois	0.22	35.87	0.45	78.68
Indiana	0.04	7.04	0.10	18.12
Iowa	0.23	52.79	0.67	143.01
Kansas	0.14	14.83	0.28	30.81
Michigan	0.01	2.45	0.04	8.56
Minnesota	0.29	30.73	0.61	66.07
Missouri	0.25	28.33	0.50	60.72
Nebraska	0.14	14.97	0.28	31.67
Ohio	0.05	9.60	0.13	24.74
Wisconsin	0.11	13.48	0.24	30.61

Figure 3-1. CRP Acreage and Corn Prices Received by Farmers, 1986-2013.

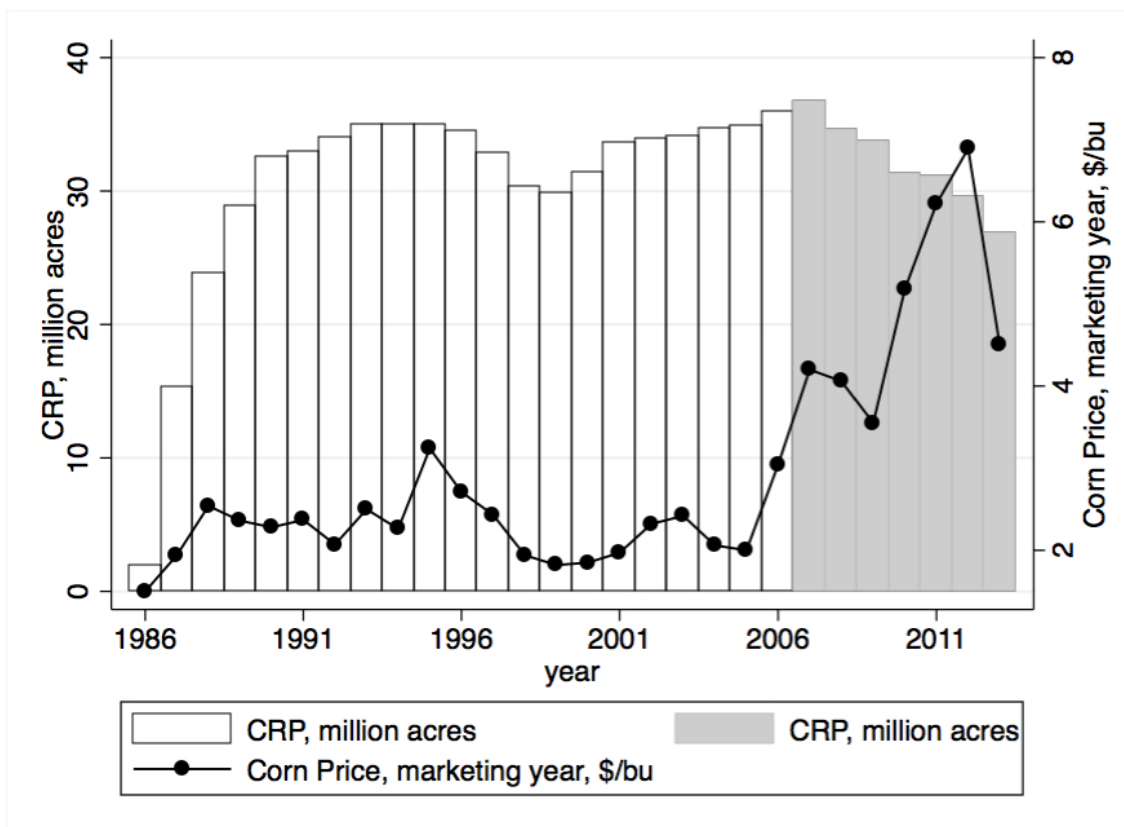


Figure 3-2. Histogram of Relative Productivity Estimates.

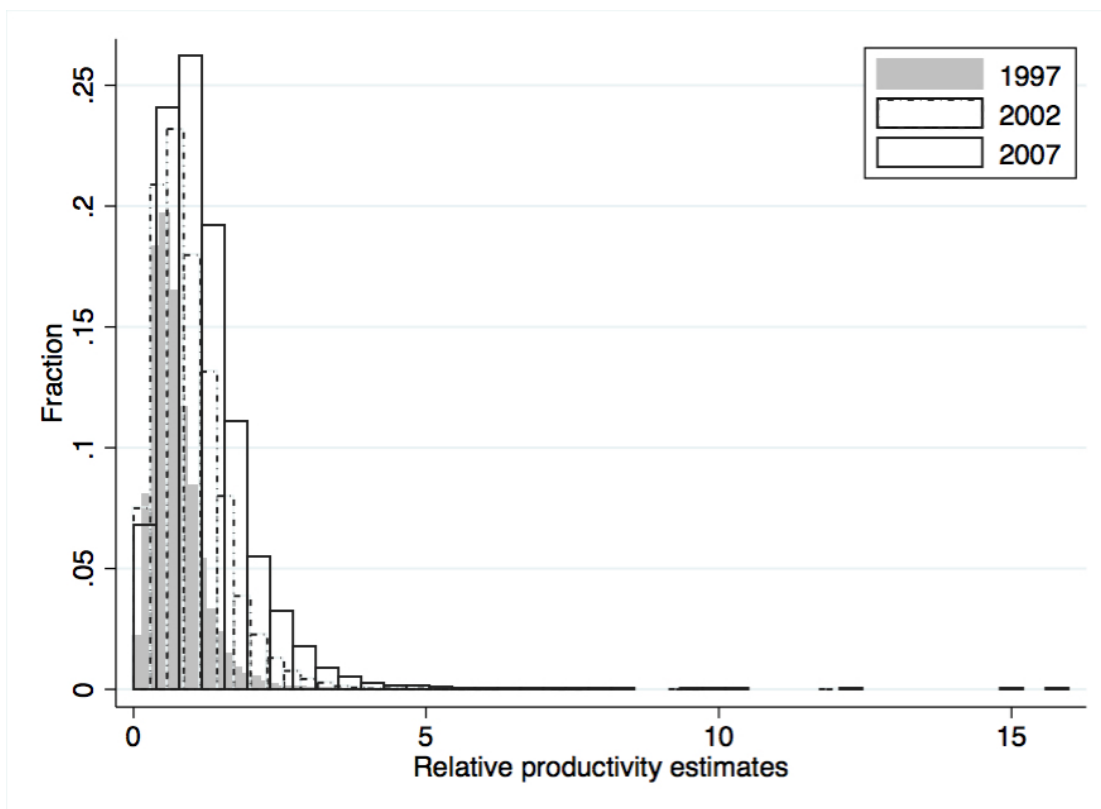


Figure 3-3. Actual Changes in CRP Acreage between 2007 and 2012 (million acres).

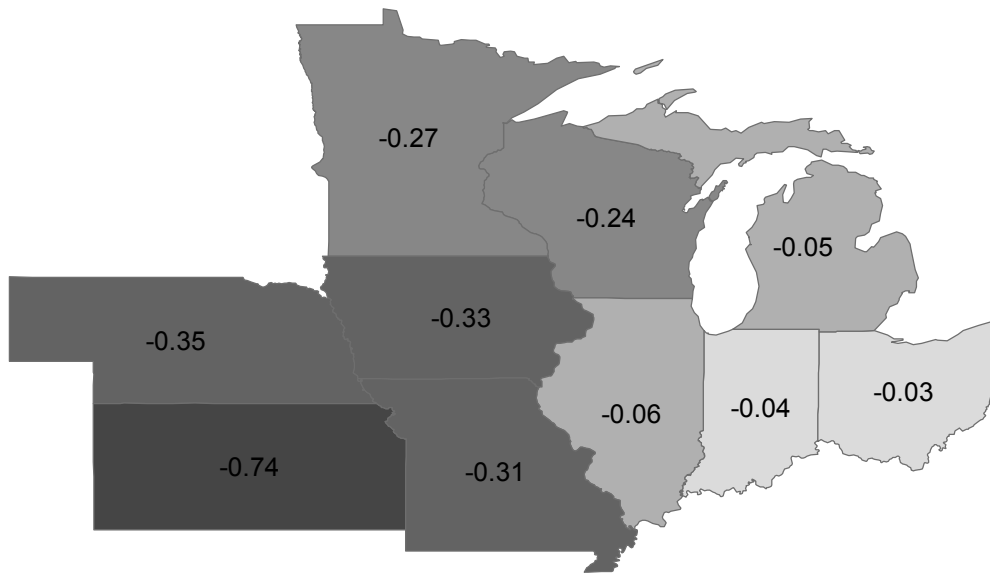


Figure 3-4. Actual Changes in CRP Costs between 2007 and 2012 (\$ millions).

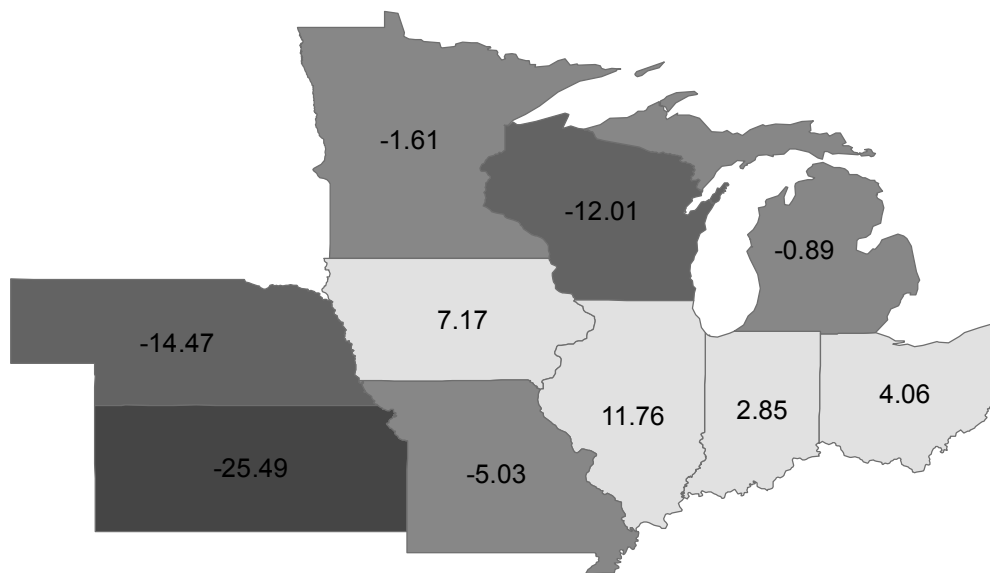


Figure 3-5. Predicted Changes in CRP Acreage between 2007 and 2012 (million acres).

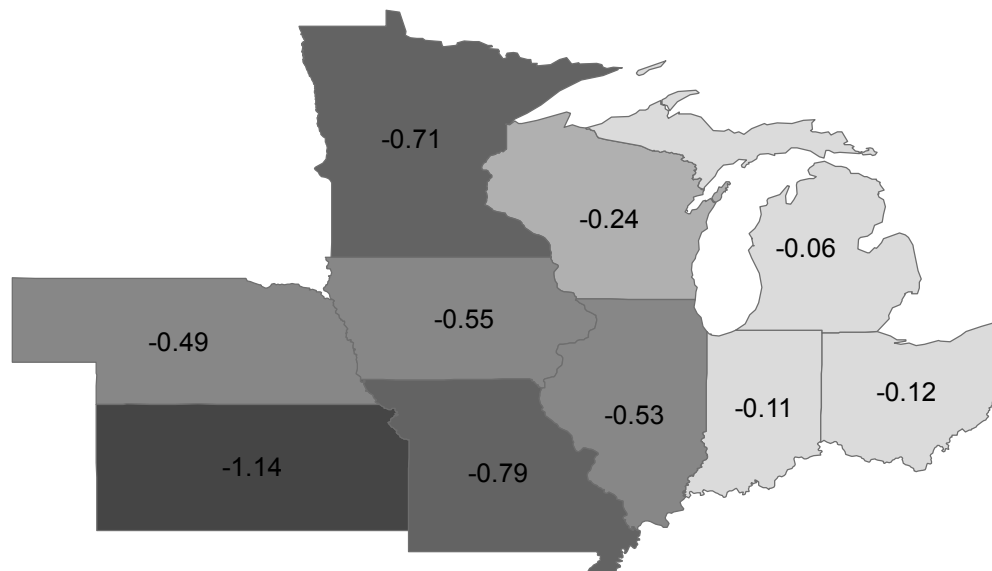


Figure 3-6. Additional CRP Acreage by 10% Increase in 2012 CRP Payment (million acres).

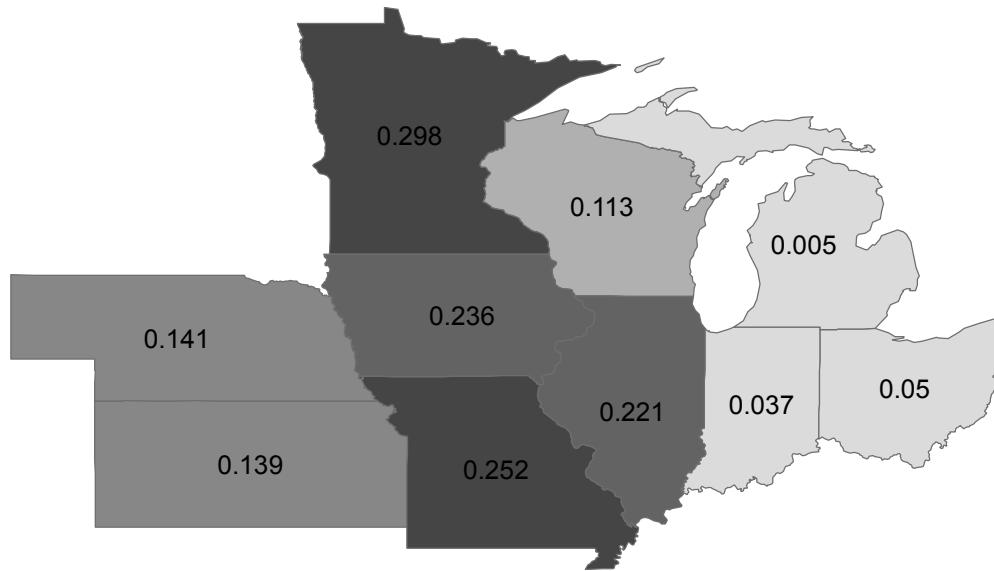


Figure 3-7. Additional CRP Costs by 10% Increase in 2012 CRP Payment (\$ millions).

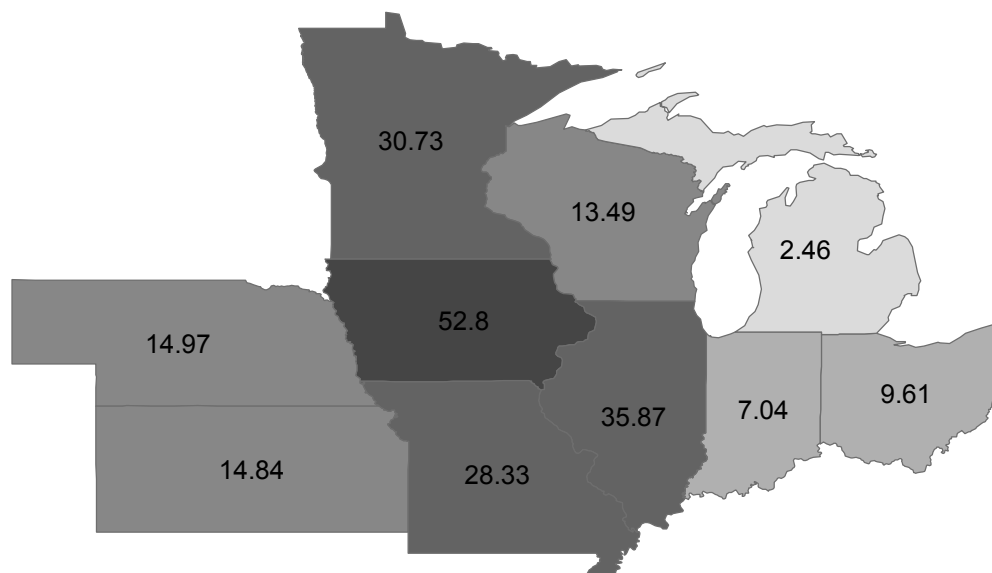


Figure 3-8. Additional CRP Acreage by 20% Increase in 2012 CRP Payment (million acres).

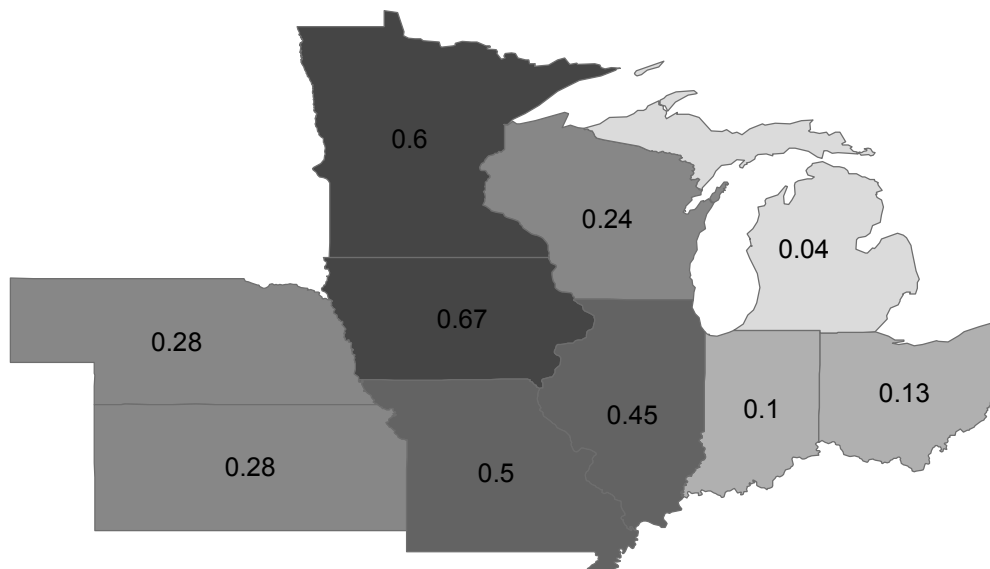
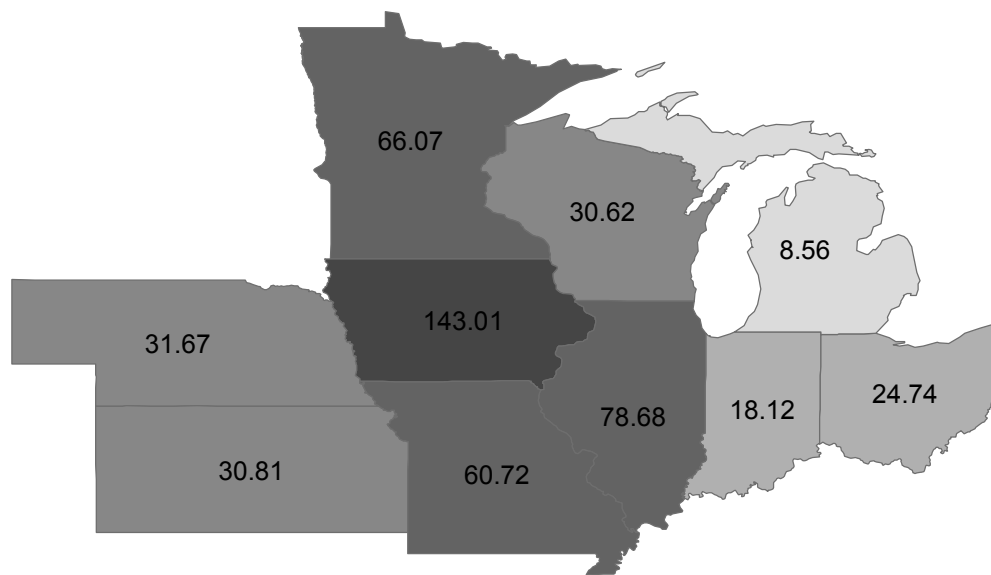


Figure 3-9. Additional CRP Costs by 20% Increase in 2012 CRP Payment (\$ millions).



## 10. Appendix

Table 3-A1. Percentage Changes in CRP Acreage.

	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	Actual changes
Illinois	-0.30	0.22	0.37	-0.91	0.76	0.61	-0.49	-0.05
Indiana	-0.33	0.24	0.41	-0.91	0.83	0.99	-0.34	-0.11
Iowa	-0.41	0.30	0.67	-0.94	1.04	1.51	-0.27	-0.16
Kansas	-0.09	0.06	0.05	-0.55	0.22	0.01	-0.34	-0.22
Michigan	-0.32	0.25	0.41	-0.88	0.84	0.78	-0.20	-0.19
Minnesota	-0.27	0.22	0.31	-0.85	0.74	0.49	-0.38	-0.14
Missouri	-0.22	0.17	0.23	-0.87	0.56	0.40	-0.49	-0.19
Nebraska	-0.16	0.12	0.15	-0.67	0.41	0.14	-0.36	-0.25
Ohio	-0.34	0.26	0.46	-0.90	0.89	0.89	-0.34	-0.07
Wisconsin	-0.34	0.27	0.37	-0.92	0.87	0.67	-0.38	-0.39

Notes:

(i) Output price increases by 10% holding other prices at the 2007 level.

(ii) Input price increases by 10% holding other prices at the 2007 level.

(iii) CRP payment increases by 10% holding other prices at the 2007 level.

(iv) Output price increases to 2012 price holding other prices at the 2007 level.

(v) Input price increases to 2012 price holding other prices at the 2007 level.

(vi) CRP payment increases to the payment in 2012 holding other prices at the 2007 level.

(vii) All variables increase to the 2012 prices.

Table 3-A2. Percentage Changes in Cropland Acreage.

	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	Actual changes
Illinois	0.03	-0.02	-0.04	0.25	-0.09	-0.07	0.06	-0.01
Indiana	0.03	-0.02	-0.03	0.19	-0.07	-0.09	0.03	0.003
Iowa	0.05	-0.03	-0.08	0.29	-0.12	-0.17	0.03	0.03
Kansas	0.03	-0.02	-0.01	0.20	-0.07	-0.005	0.11	0.06
Michigan	0.04	-0.03	-0.05	0.21	-0.11	-0.10	0.03	0.03
Minnesota	0.04	-0.03	-0.04	0.24	-0.10	-0.07	0.05	0.03
Missouri	0.05	-0.03	-0.05	0.34	-0.11	-0.08	0.10	-0.004
Nebraska	0.03	-0.02	-0.03	0.23	-0.09	-0.03	0.08	0.04
Ohio	0.04	-0.03	-0.06	0.29	-0.12	-0.12	0.05	0.01
Wisconsin	0.07	-0.05	-0.07	0.34	-0.15	-0.12	0.07	0.03

## Notes:

(i) Output price increases by 10% holding other prices at the 2007 level.

(ii) Input price increases by 10% holding other prices at the 2007 level.

(iii) CRP payment increases by 10% holding other prices at the 2007 level.

(iv) Output price increases to 2012 price holding other prices at the 2007 level.

(v) Input price increases to 2012 price holding other prices at the 2007 level.

(vi) CRP payment increases to the payment in 2012 holding other prices at the 2007 level.

(vii) All variables increase to the 2012 prices.

## Reference

- Ackerberg, D., Caves, K, and G. Frazer. 2006. Structural Identification of Production Functions. Working Paper. Department of Economics, University of California, Los Angeles.
- Aghino, P., N. Bloom, R. Blundell, R. Griffith, and R. Howitt. 2005. Competition and innovation: an inverted-U relationship. *Quarter Journal of Economics* 120(2): 701-728.
- Arellano, M. and O. Bover. 1995. Another Look At the Instrumental Variable Estimation of Error-Components Models. *Journal of Econometrics* 68:29-52.
- Barro, R.J. 1998. Notes on growth accounting. NBER Working Paper #6654. National Bureau of Economic Research.
- Blayney, D.P. 2002. The Changing Landscape of US Milk Production. USDA Statistical Bulletin Number 978.
- Blundell, R. and S. Bond. 2000. GMM Estimation with Persistent Panel Data: an Application to Production Functions. *Econometric Reviews* 19(3):321-340
- California Energy Commission. 2004. Ethanol fuel incentives applied in the U.S., reviewed from California's perspective. CEC Staff Report P600-04-001. Sacramento, CA.
- Chang, H. and R. N. Boisvert. 2009. Distinguishing between Whole-Farm vs. Partial-Farm Participation in the Conservation Reserve Program. *Land Economics* 85(1): 144-161.
- Chen, X. and M. Khanna. 2012. Explaining the reductions in US corn ethanol processing costs: testing competing hypothesis. *Energy Policy* 44: 153-159.
- Claassen R., A. Cattaneo, R. Johansson. 2008. Cost-Effective Design of Agri-Environmental Payment Programs: US Experience in Theory and Practice. *Ecological Economics* 65:737-752.
- Collier, R.J., R.B. Zimelman. 2007. Heat Stress Effects on Cattle: What We Know and What We Don't Know. 22<sup>nd</sup> Annual Southwest Nutrition and Management Conference.
- De Freitas, L.C. and S. Kaneko S. 2012. Is there a causal relation between ethanol innovation and the market characteristics of fuels in Brazil? *Ecological Economics* 74: 161-168.

- Fan, M. and S. Firestone. 2010. Firm Market Value and Production Technology. *International Journal of Industrial Organization* 28: 434-440.
- Feather, P., D. Hellerstein, and L. Hansen. 1999. Economic Valuation of Environmental Benefits and the Targeting of the Conservation Programs: The Case of the CRP. Agricultural Economics Report No.778. US Department of Agricultural Economic Research Service, Washington DC.
- Fezzi, C. and I. J. Bateman. 2011. Structural Agricultural Land Use Modeling for Spatial Agro-Environmental Policy Analysis. *American Journal of Agricultural Economics* 93(4): 1168-1188.
- Fleming, R.A. 2004. An Econometric Analysis of the Environmental Benefits Provided by the Conservation Reserve Program. *Journal of Agricultural and Applied Economics* 36:339-413.
- Gandhi, A., S. Navarro, and D. Rivers. 2013. On the Identification of Production Functions: How Heterogeneous is Productivity? Working Paper. University of Wisconsin-Madison.
- Goodwin, B.K, M.L. Vandever, and J.L. Deal. 2004. An Empirical Analysis of Acreage Effects of Participation in the Federal Crop Insurance Program. *American Journal of Agricultural Economics* 86(4): 1058-1077.
- Gould, B. 2010. Consolidation and Concentration in the US Dairy Industry. *Choices* 25 (2).
- Hellerstein, D., and S. Malcolm. 2011. The Influence of Rising Commodity Prices on the Conservation Reserve Program. US Department of Agricultural Economics Research Service, Washington DC.
- Hettinga, W.G., H.M. Junginger, S.C. Dekker, M. Hoogwijk, A.J. McAloon, and K.B. Hicks. 2009. Understanding the reductions in US corn ethanol production costs: An experience curve approach. *Energy Policy* 37: 190-203.
- Jacobs, K., W. Thurman, and M. Marra. 2014. The Effect of Conservation Priority Areas on Bidding Behavior in the Conservation Reserve Program. *Land Economics* 90 (1): 1-25.
- Jacobson, S. 2014. Temporal Spillovers in Land Conservation. *Journal of Economic Behavior & Organization*.
- Jaffe, A.B. 1986. Technological opportunity and spillovers of R&D: Evidence from firm's patents, profits, and market value. *The American Economic Review* 76(5): 984-1001.

- Jaffe, A.B. and M. Trajtenberg. 1996. Flows of knowledge from universities and federal laboratories: Modeling the flow of patent citations over time and across institutional and geographic boundaries. *Proceedings of the National Academy of Sciences* 93: 12671-12677.
- Johnstone N., I. Hascic, and D. Popp. 2008. Renewable energy policies and technological innovation: Evidence based on patent counts. NBER Working Paper #13760. National Bureau of Economic Research.
- Karmarkar-Deshmukh, R. and C.E. Pray. 2009. Private Sector Innovation in Biofuels in the United States: Induced by Prices or Policies? *AgBioForum* 12: 141-148.
- Kasahara, H. and J. Rodrigue. 2008. Does the Use of Imported Intermediates Increase Productivity? Plant-Level Evidence. *Journal of Development Economics*. 87(1):106-118.
- Kirwan, B., R. N. Lubowski, and M. J. Roberts. 2005. How Cost-Effective are Land Retirement Auctions? Estimating the Difference between Payment and Willingness to Accept in the Conservation Reserve Program. *American Journal of Agricultural Economics* 87:1239-1247.
- Levinsohn, J. and A. Petrin. 2003. Estimating Production Functions Using Inputs to Control for Unobservables. *Review of Economic Studies* 70: 317-340.
- Lubowski, N. Ruben, Shawn Bucholtz, Roger Claassen, Michael J. Roberts, Joseph C. Cooper, Anna Gueorguieva, and Robert Johansson. 2006. Environmental Effects of Agricultural Land-Use Change: The Role of Economics and Policy. US Department of Agricultural Economics Research Service, Washington DC.
- MacDonald, J.M., E.J. O'Donoghue, W.D. McBride, R.F. Nehring, C.L. Sandretto, and R. Mosheim. 2007. Profits, Costs, and the Changing Structure of Dairy Farming. USDA Electronic Information Bulletin Number 47.
- Melitz, M. and S. Polanec. 2012. Dynamic Olley-Pakes Productivity Decomposition with Entry and Exit. *RAND Journal of Economics*, forthcoming.
- Olley, S.G. and A. Pakes. 1996. The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica* 64(6): 1263-1297.
- Petrick, M. and M. Kloss. 2013. Identifying Factor Productivity from Micro-data: The Case of EU Agriculture. Factor Markets Working Paper. Centre for European Policy Studies.
- Popp, D. 2002. Induced innovation and energy prices. *American Economic Review* 92(1): 160-180.

- Popp, D. 2010. Innovation and climate policy. NBER Working Paper #15673. National Bureau of Economic Research.
- Renewable Fuel Association (RFA). 2012. Ethanol industry statistics. Retrieved online at <http://www.ethanolrfa.org/pages/statistics>. (Accessed December 18, 2012).
- Ribaudo, M.O., D. Colacicco, L. L. Langner, S.Piper, and G.D.Schaible. 1990. Natural Resources and User Benefit from the Conservation Reserve Program. Agricultural Economics Report No.627. US Department of Agriculture Economics Research Service, Washington DC.
- Richard, B., and S. Bond. 2000. GMM Estimation with Persistent Panel Data: An Application to Production Functions. *Econometric Reviews* 19(3): 321-340.
- Schlenker, W. and M.J. Roberts. 2009. Nonlinear Temperature Effects Indicate Severe Damages to U.S. Crop Yields under Climate Change. *Proceedings of the National Academy of Sciences of the United States of America* 106(37): 15594-15598.
- Sieg, H. and J. Zhang. 2012. The Importance of Managerial Capacity in Fundraising: Evidence from Land Conservation Charities. *International Journal of Industrial Organization* 30: 724-734.
- Smith, R.B. 1995. The Conservation Reserve Program as a Least-Cost Land Retirement Mechanism. *American Journal of Agricultural Economics* 77(1):93-105.
- Stata Corporation. 2001. *STATA 7 reference manual*. College Station, TX: Stata Press.
- Stubbs, M. 2014. Conservation Reserve Program (CRP): Status and Issues. Congressional Research Service.
- Sullivan, P., D. Hellerstein, L. Hansen, R. Johansson, S. Koeing, R. Lubowski, W. McBride, D. McGranahan, M. Roberts, S. Vogel, and S. Bucholtz. 2004. The Conservation Reserve Program: Economic Implications for Rural America. Agricultural Economic Report No.834. US Department of Agriculture Economics Research Service, Washington DC.
- Sumner, D.A. and C.A. Wolf. 2002. Diversification, Vertical Integration, and the Regional Pattern of Dairy Farm Size. *Review of Agricultural Economics* 24(2): 442-457.
- Verdolini, E. and M. Galeotti. 2011. At home and abroad: An empirical analysis of innovation and diffusion in energy technologies. *Journal of Environmental Economics and Management* 61: 119-134.

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica* 48 (4): 817–838.

Wooldridge, J.M. 2009. On Estimating Firm-Level Production Functions Using Proxy Variables to Control for Unobservables. *Economic Letters* 104(3):112-114.