

**From the holdings of the Department of Special Collections , Princeton University Library,
Princeton, NJ**

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyright law. We do not charge any permission or usage fees for the publication of images of material in our collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff. For more information regarding copyright, credit, and citation please review the information on the Copyright, Credit and Citation Guidelines page of our website.

<https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines>. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. <https://library.princeton.edu/special-collections/ask-us>

Doctoral Dissertations, 2011-2026

Economics

2013

Essays on the Estimation of Food Supply and Demand

Scott, Paul Thomas

Princeton, NJ : Princeton University

<http://arks.princeton.edu/ark:/88435/dsp01x633f115v>

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. <https://www.copyright.gov/title17/title17.pdf>

**ESSAYS ON THE ESTIMATION OF
FOOD SUPPLY AND DEMAND**

PAUL THOMAS SCOTT

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF
ECONOMICS
ADVISOR: JAN DE LOECKER

SEPTEMBER 2013

© Copyright by Paul Thomas Scott, 2013. All rights reserved.

Acknowledgments

My advisor, Jan De Loecker, has provided good guidance, insightful criticism, and steady encouragement since I first started considering dissertation topics. He somehow always found time when I needed it, even while in absentia, starting a family, and on the tenure track. I cannot thank him enough.

I thank Bo Honoré for all his support over the years. When I was stuck on a problem, he had a solution. When I was confused, he had a perfectly clear explanation.

Steve Berry gave me confidence in my choice of research topics when I needed it most, and brought me to Yale for many incredibly productive conversations. I am grateful to Myrto Kalouptsi and Sylvain Chassang for numerous helpful conversations about research and guidance in navigating the job market. Tim Searchinger has been a constant source of knowledge and inspiration on the topic of land use change. I thank Ariel Pakes for sponsoring a visiting semester at Harvard, where I learned much and performed some of my thesis research.

This dissertation was improved by comments and suggestions from Allan Collard-Wexler, Joseph Hotz, Robert Miller, Eduardo Morales, Andriy Norets, Sergio Pastorello, Penny Goldberg, Nathan Yang, and various lunch workshop and seminar participants. I thank Wolfram Schlenker for sharing data with me, and I thank Patrick Willis for extensive help in understanding the Cropland Data Layer.

The Center for Economic Policy Studies and Econometrics Research Program at Princeton generously provided research funding. I am deeply grateful for support from Resources for the Future's Joseph L. Fisher Dissertation Fellowship.

My dissertation benefited tremendously from several people in the agricultural industry and at regulatory agencies who generously shared their knowledge of the workings of food markets with me. I am especially grateful to Mike Fischer, Scott Hillman, Jeff Hockett, Pat Kirby, Don Shippelhouse, Don Stranik, Jeff Wilbur, and Glenn Yost for generously sharing their time.

Finally, I thank my parents. They will always be my greatest influences and biggest supporters.

Abstract

I develop and apply structural methods for estimating supply and demand responses of food products and agricultural commodities.

Chapter 1 adapts recent developments in dynamic discrete choice econometrics to estimate a model of land use change. The empirical approach is relatively easy to implement, relying on linear regressions. I apply the empirical approach to US crop supply using satellite sensor data. Taking dynamics into account leads to considerably larger long-run elasticities, and suggests that the land-use impacts of biofuels production are larger than previous studies have found.

Chapter 2 considers the intensive margin in crop supply responses. I derive and implement an indirect approach to estimating yield-price elasticities which is considerably more precise than the standard approach. I find that yields for US corn, soybeans, and wheat respond very little to short-run price variation, suggesting that extensive crop supply responses (i.e., land use change), is the dominant component of crop supply responses, at least in the short run within the US.

Chapter 3 analyzes the fluid milk market, including non-dairy milks like soy milk, as a case study in understanding the potential for greenhouse gas mitigation through dietary change. I estimate a random coefficients model of demand for milk products using retail data, and find that taking unobservable heterogeneity into account suggests that there is substantially less substitution between milk and soy milk in response to price changes. A static demand model cannot explain long-term changes in milk consumption patterns, for soy milk consumption rose dramatically in the early 2000's without any substantial changes in relative prices.

Contents

Chapter 1. Dynamic Discrete Choice Estimation of Agricultural Land Use	1
1.1. Introduction	2
1.2. Empirical framework	6
1.3. Data and measurement	14
1.4. Estimation	19
1.5. Results	26
1.6. Implications for the US biofuels mandate	28
1.7. Conclusion	30
1.8. Technical Appendix	30
1.9. Data appendix	35
1.10. Sensitivity analysis	38
References	51
Chapter 2. Indirect Estimation of Yield-Price Elasticities	55
2.1. Introduction	56
2.2. Theory	58
2.3. Data and estimation	62
2.4. Results	66
2.5. Conclusion	67
References	67
Chapter 3. Climate-Saving Dietary Change: The Case of Soy Milk	69
3.1. Introduction	70
3.2. Model and estimation	71
3.3. Results	73
3.4. Milk consumption over time	77
References	78

CHAPTER 1

Dynamic Discrete Choice Estimation of Agricultural Land Use

1.1. Introduction

The effects of many controversial policies related to greenhouse gas mitigation, ecological destruction, and agricultural policy depend crucially on how land use responds to economic changes. Land use change is a fundamentally dynamic process, often involving switching costs (e.g., clearing forest to plant crops) and sometimes involving switching benefits (e.g., nutrient and pest management provide strong incentives for rotation between crops such as corn and soybeans in some cases). Nevertheless, empirical studies of land use change typically treat landowners as static decision makers. In contrast, I formulate and estimate a flexible and easily implemented empirical approach for analyzing land use based on a model of dynamically optimizing landowners.

While land use change is the central concern in many policies involving payments for environmental and ecosystem services (e.g., the UN's REDD program, the Conservation Reserve Program in the US, and agri-environmental programs in the EU), land use change also plays a prominent role in policies which do not target land use directly. Broadly, changes in the area of cultivated land are an important aspect of agricultural supply responses, so any question which depends on supply elasticities for agricultural commodities can be said to depend, in part, on land use responses. Conversely, any policy affecting agricultural markets can have indirect land use effects.

Indirect land use change has become a primary concern in evaluating biofuels regulation. Many governments mandate that some portion of their countries' fuel supplies come from biofuels.¹ The primary feedstocks for biofuel production around the world are crops (especially corn, sugarcane, and various oil crops), so biofuels mandates effectively increase crop demand. In the US, 35-40% of corn production has been used to produce ethanol in recent years (US EPA, 2011).² Properly evaluating the equilibrium effects of the increased demand created by biofuels mandates requires an understanding of land use elasticities. On the one hand, if cropland use responds little to changes in crop prices, elevated crop demand will lead to elevated crop prices, and decreased use of crops for other purposes (such as direct human consumption and animal feed). On the other hand, a more elastic crop acreage response would mitigate the effect on food prices, but result in higher environmental costs through indirect land use change. The most influential recent research on the equilibrium effects of biofuels relies on static models, both in estimating supply and demand

¹In the US, bioethanol subsidies expired in January, 2012, but mandated levels of biofuel use remain in effect, and the growing demands of the mandate continue to be met mostly by corn ethanol (US EPA, 2011, 2012). In 2011, the mandate of 13.95 billion gallons of biofuels represents about 9% of US gasoline consumption. The EU mandates biofuels use on similar levels (Flach et al., 2012). Brazil's mandate is larger in relative terms, with 18-25% of Brazilian gasoline blends coming from bioethanol (Barros, 2012). China, India, and many other countries have biofuels mandates.

²There are now widespread efforts to develop cost-effective biofuel production strategies for feedstocks that don't compete with sources of human feed or animal food (e.g., algae). However, conventional biofuels remain substantially less costly to produce for the time being.

elasticities (Roberts and Schlenker, 2013) and in simulating equilibrium responses (Searchinger et al., 2008; Tyner et al., 2010).

I estimate a dynamic discrete choice model of cropland use in the United States, finding a long-run elasticity of crop acreage with respect to crop prices in the neighborhood of 0.4. This elasticity is more than five times larger than static elasticities estimated using the same data.

To give the results some context, I revisit Roberts and Schlenker’s (2013) assessment of the effects of the US biofuels mandate, finding that my long-run supply elasticity implies a 35% larger land use effect and a 70% smaller price increase (in the long run).

I employ a novel empirical approach which combines three powerful ideas. First, the model may be estimated using a linear regression, derived using the Hotz-Miller inversion (Hotz and Miller, 1993, Proposition 1).³ Second, while the estimation of dynamic discrete choice models typically requires a model of how all state variables evolve, this burden can be relaxed when the agents in the model have a negligible impact on some market-level state variables – in this case, the uncertainty in market-level state variables can be integrated out into a well-behaved regression error term.⁴ Finally, the Expectation-Maximization (EM) algorithm can be used to estimate the model with unobservable heterogeneity, following Arcidiacono and Miller (2011).

My empirical strategy is broadly applicable to dynamic discrete choice settings with price-taking agents. In the context of land use, the decisions of individual landowners plausibly have a negligible effect on prices in agricultural commodity markets. My strategy could be applied to other dynamic discrete choice applications with less restrictive assumptions than those used by previous studies; for instance, my approach could be used to estimate demand for durable goods without having to assume that consumers have perfect foresight (Conlon, 2010) or even that prices evolve according to a particular sort of process (Hendel and Nevo, 2006; Gowrisankaran and Rysman, 2011).

While static models remain common in empirical work on land use (Chomitz and Gray, 1996; Lin et al., 2000; Deininger and Minten, 2002; Chomitz and Thomas, 2003; Fezzi and Bateman, 2011; Souza-Rodrigues, 2012), some strategies incorporate state dependence without forward-looking dynamics (Claassen and Tegene, 1999; Lubowski, 2002; Munroe et al., 2004; Wu et al., 2004; Lubowski et al., 2006, 2008). Existing studies on land use which account for forward-looking behavior are rare and largely confined to models of irreversible

³Although it has been poorly documented, it is generally possible to construct linear estimators for dynamic discrete choice models. For example, Pesendorfer and Schmidt-Dengler (2008) construct a representation of equilibrium conditions which is linear in parameters (Lemma 1 in their appendix).

⁴This procedure amounts to the construction of an Euler equation for a dynamic discrete choice problem. As the term is used in economics, Euler equations typically involve differentials with respect to continuous choice variables; naturally, the discrete choice analog involves differences across actions. As in Hall (1978), errors in agent’s expectations about continuation values will be mean uncorrelated with anything in the contemporaneous information set. Although its general applicability in single-agent dynamic discrete choice settings has largely been overlooked, differenced Euler equation constructions are implicitly used by Altug and Miller (1998) and Murphy (2012).

decisions (Irwin and Bockstael, 2002; Vance and Geoghegan, 2002; Murphy, 2012). As my results suggest, models of myopic landowner behavior are likely to understate long-run land use responses. Intuitively, landowners may respond more to long-run changes in the process governing price changes than they respond to the price variation in the data. However, a myopic model only captures how land use responds to the price variation in the data. In contrast, my empirical model is based on dynamically optimizing landowners with rational expectations.

An important practical aspect of my empirical strategy is that it does not require the econometrician to estimate (or make restrictive functional form assumptions about) the process governing the evolution of input prices, output prices, stocks, and other market-level state variables. In contrast, standard approaches to dynamic discrete choice estimation such as Rust (1987), Hotz and Miller (1993), and Aguirregabiria and Mira (2002) require models of how all state variables evolve. In the context of land use change, market-level state variables should be modeled at the local level, but local information on variables like prices and stocks in storage is generally very limited. Moreover, even if all relevant variables were observed at the local level, modeling their evolution may present an infeasible task, for agricultural commodity markets are largely integrated – in principle, the set of state variables for any locality would include prices and stocks for every other locality around the world. Thus, precisely modeling the evolution of market-level state variables is infeasible for reasons of tractability as well as data limitations. By avoiding the need for such a model, my empirical approach avoids the need for incredible simplifying assumptions about the evolution of market-level state variables. Instead, my regression equation construction relies on the assumption that landowners have rational expectations. In other words, the agents in the model must have a model of how state variables evolve, but the econometrician doesn't need one.

In empirical studies on land use change, researchers typically project local variables using aggregate data together with spatial information (e.g., projecting local prices using prices at commodity centers and distances from those centers as in De Pinto and Nelson (2009) and Souza-Rodrigues (2012)). However, such projections are fundamentally imprecise, and the resulting projection errors could be treated as unobservable supply shocks. For example, it would be tempting to assume that farmers throughout in the US Corn Belt expect to receive prices which correspond to prices in Chicago and the costs of shipping corn from their farm to Chicago. However, farmers in some areas might receive slightly higher prices in some years by selling their corn to local livestock operations. Such pricing patterns are likely to be influenced by factors which are highly correlated over time but imperfectly persistent (such as feed demand from nearby livestock operations and local grain storage), meaning that projection errors should be included as unobservable shocks which may be serially correlated – i.e., such errors are likely to violate the conditional independence assumptions typically placed on idiosyncratic error terms in dynamic discrete choice models.

Fortunately, my approach can accommodate unobservable market-level shocks, even if they are serially correlated. Such shocks become error terms in a regression equation, and while consistent estimation of the model requires exclusion restrictions on the unobservable shocks, the unobservable shocks may have an arbitrary correlation structure subject to those exclusion restrictions.

Unobservable heterogeneity is another practical challenge which is difficult to avoid when modeling land use. While detailed spatial information on soil and weather characteristics provides a wealth of information about field-level heterogeneity, fields may differ on such a multitude of characteristics that it may be infeasible to account for every payoff-relevant dimension of heterogeneity.⁵ When field-level characteristics cannot be quantified completely, ignoring unobservable heterogeneity can lead to biased estimates (e.g., when ignored, persistent unobservable heterogeneity may exaggerate switching costs). To estimate my dynamic model with unobservable heterogeneity, I follow Arcidiacono and Miller (2011) in using the EM algorithm to estimate a mixture model of conditional choice probabilities.

The main data set used for estimation is a unique panel of land use outcomes covering the contiguous United States in recent years. In terms of covariates, a measure of expected returns is the only input to the model.⁶ However, specifying the choice model in this way calls for careful measurement of expected returns (with expectations taken at the time when land use choices are made, i.e., during planting season). I construct expected returns based on state-level price forecasts, county-level yield forecasts, aggregate cost data, and government payment rates.

Typically, the lack of cross-sectional variation in agricultural commodity prices limits identification of complex models of agricultural supply. Although output prices vary little across locations, there are persistent cross-sectional differences in yields which I leverage to construct cross-sectional variation in (expected) returns. My yield forecasts, based on historical county-level yields and detailed weather data, perform extremely well in validation tests. My model of expected yields relies on Schlenker and Roberts's (2009) work on nonlinear temperature effects in crop yields (and their weather data).

In Section 1.2, I lay out a binary choice model of land use and derive a regression equation. In Section 1.3, I describe data sources, the construction of the land use panel data set, and the measurement of expected returns. Further estimation details are treated in Section 1.4, including the extension to unobservable heterogeneity. Section 1.5 presents the results, Section 1.6 considers implications for biofuels policy, and Section 1.7 concludes.

⁵For example, local economic characteristics such as proximity to processing, storage, and input manufacturing facilities may be almost as important as soil characteristics, but they are much harder to quantify at a fine level of spatial resolution. One might argue that these things could be measured, too, but the end to this argument is nowhere in sight. To cite another example, it would be difficult to capture *everything* that affects a field's suitability for cultivation even with extremely detailed terrain data, for a field's susceptibility to flooding can be influenced by very small variations in local topography.

⁶While returns may be the primary consideration in most private land-use decisions, my model is arguably less applicable to government-managed land. Accordingly, protected land is excluded from the sample.

1.2. Empirical framework

This section describes a flexible model of land use with dynamically optimizing agents. To simplify the exposition, I consider fields of a homogeneous type. That is, fields may differ due to differences in the history of actions they take, and due to idiosyncratic shocks, but they are otherwise similar – e.g., they should face similar prices and weather patterns in expectation. In Section 1.4, I explain how I estimate the model with different observable types and with a mixture of unobservable types.

1.2.1. Model. Field owners act to maximize the expected discounted profits from their fields. Each year, during planting season, field owners decide whether to plant crops in their fields or not. Formally, the choice set is $\mathbf{J} = \{crops, other\}$. Let j_{it} denote the land use of field i in year t .⁷

There are two types of state variables in the model. First, the *field state* (k_{it}) represents characteristics intrinsic to field i at time t . For example, field states may represent soil nutrient levels, the state of the terrain, or enrollment in a government program. Let the set of field states be denoted by $\mathbf{K}$, which is assumed to be discrete.

The other state variable is the information set or *market state* (ω_t), which includes all the information necessary to determine current expected returns for a given field state (e.g., futures prices, input costs, inventories) as well as information which is relevant in predicting future market states (e.g., demand and policy conditions). The current market state is known to all field managers but not fully observable to the econometrician. Let Ω denote the set of possible market states.

During planting season, returns are uncertain even in the current year. For example, weather is an intrinsic source of randomness in crop yields, and input and output prices fluctuate over the course of the growing season (a stark example is the US drought during the summer of 2012, which caused yields in the Midwest to fall far below expectations and prices to rise well above expectations). I assume field managers are risk neutral so that expected returns are all that is needed to model their decisions.

If field i is in state k at time t , the expected payoffs to land use j are

$$(1.1) \quad \pi(j, k, \omega_t, \nu_{it}) = \alpha_0(j, k) + \alpha_R R_j(\omega_t) + \xi_{jk}(\omega_t) + \nu_{jit}$$

where $R_j(\omega_t)$ is an observable (to the econometrician) component of expected returns, $\xi_{jk}(\omega_t)$ is an unobservable aggregate shock to expected returns, ν_{jit} is an idiosyncratic shock, and $\alpha = \left(\alpha_R, \{\alpha_0(j, k)\}_{j \in \mathbf{J}, k \in \mathbf{K}} \right)$ is the vector of parameters to be estimated. For the purposes of this section, take for granted that $R_j(\omega_t)$ is available data; Section 1.3 explains how I construct the variable.

⁷The empirical approach I describe can be generalized to larger discrete choice sets. See Appendix 1.8.6.

The inclusion of the unobservable shock ξ in the profit equation (1.1) can be motivated by the limitations of data availability, capturing the imperfect measurement of returns because local values of input and output prices are unavailable. When it comes to estimation, I don't construct a measure of observable returns for non-cropland, so the unobservable shocks actually capture all of the variation in non-cropland returns.

Dynamic incentives come from the dependence of the intercept term $\alpha_0(j, k)$ on the field state k . While the assumption that the field state shifts only the intercept term is restrictive (i.e., field states can affect switching costs but not productivity), it is relatively common in the literature (Claassen and Tegene, 1999; Munroe et al., 2004; Wu et al., 2004), so I adopt it as a starting point.⁸

For the purposes of my estimation strategy, the important difference between field states and market states is that field states must be observable and the econometrician must know the process governing their evolution (or be able to estimate it). In contrast, the econometrician does not need to observe all market-level state variables, and no functional form assumptions are necessary on the process governing the evolution of the market state.

ASSUMPTION 1. (Small fields, no externalities) *The market state evolves according to a Markov process which is unaffected by changing the land use in any single field; i.e., the conditional distribution of ω_{t+1} satisfies $G(\omega_{t+1}|\omega_t, j_{it} = j) = G(\omega_{t+1}|\omega_t)$ for all i and j .*

The assumption of small fields implies that, although the process governing the evolution of market states is endogenous in general, it may be regarded as exogenous by a small agent in a competitive equilibrium. Given that agricultural commodity markets are highly integrated, and changing an individual field's usage plausibly has a negligible effect on prices and other aggregate variables, the assumption of price-taking fields provides a very plausible approximation of reality.⁹ While the assumption that there are no externalities across fields is restrictive in general, effects across fields plausibly play little role in the decision of whether to plant crops.¹⁰

⁸Lubowski (2002), Lubowski et al. (2006), and Lubowski et al. (2008) allow field states (lagged land use decisions in their case) to affect coefficients on covariates as well as intercept terms. This is straightforward in their model, for without forward-looking behavior, they can effectively estimate a different model for each field state.

⁹The 2007 US Census of Agriculture reports that there were 310 million acres of harvested cropland in the US spread over 1.3 million farms. Furthermore, 170 million acres of cropland were spread across farms of under 5000 acres in size, and The Land Report magazine claims that no American individual or business owns over 2.2 million acres of land (see the 2012 Land Report 100). Thus, ownership of cropland in the United States is highly unconcentrated.

¹⁰Anecdotal evidence suggests that production complementarities across fields in different type of crops are important for some farmers – for example, while there are powerful dynamic considerations driving the corn-soybeans rotation (pest and soil nutrient management), another reason why farmers might keep half their land in soybeans and half in corn (switching the two from year to year) is that the two crops can be planted and harvested at slightly different times, potentially saving on labor costs. In some areas, rotation between cropland and fallow land may similarly be driving by both dynamic considerations and economies of space. Since I rule out the externalities across fields, these economies of space may be absorbed as dynamic effects.

Given Assumption 1, it is without loss of generality to assume that each landowner manages a single field, so i can be used to refer to an agent or the field she manages. Without market power or externalities across fields, maximizing the individual profits of several fields is equivalent to maximizing their joint profits.

I assume that field states are a deterministic function of past land use, and that planting crops is a *renewal action* always leading to the same field state.¹¹ However, a field's state can evolve when it remains in non-cropland, potentially capturing several effects. First, if the outside option is leaving the field idle, then the land might slowly revert to natural terrain, and the costs of switching back to crops might increase during the reversion process. Alternatively, in some areas leaving fields unplanted (fallow) is an important part of a dynamic management process, much like crop rotation – in this case the costs of planting crops might fall after the land is left fallow for a year.

Formally, I let the field state denote the number of years since crops were last planted in the field, up to some limit $\bar{k}$, implying that the set of possible field states is $\mathbf{K} = \{0, 1, \dots, \bar{k}\}$. Formally, the state transition process is

$$(1.2) \quad \kappa^+(j, k) = \begin{cases} 0 & \text{if } j = \text{crops} \\ \min\{k + 1, \bar{k}\} & \text{if } j = \text{other}. \end{cases}$$

Thus, if crops were planted in field i in year $t - 1$, then $k_{i,t} = 0$. If that same field is then used for non-cropland in year t , then $k_{i,t+1} = 1$. If the field continues to be used for non-cropland indefinitely, then $k_{i,t+s} = \bar{k}$ for $s \geq \bar{k}$. The special case with $\bar{k} = 1$ corresponds to a model in which the profit equation is affected only by the previous land use (as in Claassen and Tegene (1999), Lubowski (2002), Wu et al. (2004), Lubowski et al. (2006), and Lubowski et al. (2008)).

Next, I adopt the standard logit model assumption.

ASSUMPTION 2. (Conditionally independent logit errors) *Conditional on ω_t and k_{it} , ν_{jit} is identically and independently distributed across i , j , and t with a type 1 extreme value distribution.*

Assumption 2 implies that differences in idiosyncratic error terms have a logistic distribution, resulting in convenient expressions for the value functions and conditional choice probabilities. Without loss of generality, I normalize the variance of ν_{jit} to $\frac{\pi^2}{6}$, implying that the distribution function is $F(\nu_{jit}) = \exp(-\exp(-\nu_{jit}))$.¹²

¹¹A renewal action is a special case of finite dependence. See Arcidiacono and Miller (2011) for a formal definition of finite dependence, and see Arcidiacono and Ellickson (2011) for an overview of how finite dependence leads to simple estimation approaches.

¹²The sensitivity parameter α_R can be seen as a result of this normalization – i.e., α_R is inversely proportional to the variance of the idiosyncratic errors when they are measured in the same units as returns.

I now consider a field owner's dynamic optimization problem. Let β represent a common discount factor. Field owner i 's value function is defined as follows:

$$(1.3) \quad V(k_{it}, \omega_t, \nu_{it}) \equiv \max_{j^*} E \left(\sum_{s \geq t}^{\infty} \beta^{s-t} \pi(j^*(k_{is}, \omega_s, \nu_{is}), k_{is}, \omega_s, \nu_{is}) | k_{it}, \omega_t, \nu_{it} \right).$$

where the maximization is over all policy functions $j^* : \Omega \times \mathbf{K} \times \mathbb{R}^J \rightarrow \mathbf{J}$.

With respect to the process governing the evolution of the market state ω_t , my empirical approach is very flexible. The process must be well behaved enough for the value function to exist,¹³ Assumption 1 must hold, and estimation will require identifying assumptions on the unobservable shocks ξ . However, the process governing the evolution of ω_t does not have to be modeled explicitly or estimated. Hereafter, I will mainly use t subscripts on functions and variables which depend on ω_t ; e.g., $V_t(k_{it}, \nu_{it}) = V(k_{it}, \omega_t, \nu_{it})$, and $R_{jt} = R_j(\omega_t)$.

The *ex ante value function* is the expectation of the value function before the realization of idiosyncratic errors:

$$(1.4) \quad \bar{V}_t(k) \equiv \int \dots \int V_t(k, (\nu_1, \dots, \nu_J)) dF(\nu_1) \dots dF(\nu_J).$$

The *conditional value function* represents the expected discounted returns conditional on an action, but before the realization of ν_{it} :

$$(1.5) \quad \delta_t(j, k) \equiv \bar{\pi}_t(j, k) + \beta E_t [\bar{V}_{t+1}(\kappa^+(j, k))].$$

where $\bar{\pi}_t(j, k) \equiv \pi(j, k, \omega_t, 0)$. Note that the expectation of the value function at $t+1$ does not need to be conditioned on j because of Assumption 1.

Next, Assumption 2 implies a simple expression for *conditional choice probabilities*. Defining $p_t(j, k) \equiv Pr(j_{it} = j | k_{it} = k, \omega_t)$,

$$(1.6) \quad p_t(j, k) = \frac{\exp(\delta_t(j, k))}{\sum_{j' \in J} \exp(\delta_t(j', k))}.$$

The logit errors assumption also implies a convenient expression for the mean value function:

$$(1.7) \quad \bar{V}_t(k) \equiv \ln \left(\sum_{j \in J} \exp(\delta_t(j, k)) \right) + \gamma$$

where γ is Euler's gamma.

¹³See Bhattacharya and Majumdar (1989) for regularity conditions on G which guarantee the existence of the value function.

1.2.2. Deriving a regression equation. In this section, I derive a linear regression equation for the model presented above. Appendix 1.8.6 explains how this derivation can be generalized to larger choice sets and different assumptions on the idiosyncratic error terms.

Pesendorfer and Schmidt-Dengler (2008) show that it is possible quite generally to construct representations of equilibrium conditions in dynamic discrete models which are linear in parameters (see Lemma 1 in their appendix). However, their construction relies on an explicit model of how all state variables evolve. The distinguishing aspect of the construction I present here is that it avoids this requirement. Given that agents are small and have rational expectations, market-level state variables can be integrated out into an expectational error term which is a “true regression disturbance,” uncorrelated with any variables in ω_t (as in Hall (1978)).

The derivation amounts to constructing an Euler equation out of conditional choice probabilities. Traditionally, Euler equations in economics involve marginal effect changes in a continuous choice variable. Naturally, the discrete choice analog of an Euler equation requires that we consider two distinct alternative actions, rather than a marginal change, but in both cases, the Euler equation equates the benefits of a change today with the costs of a compensating change tomorrow.

I begin with the Hotz-Miller inversion, which states that there is an invertible mapping between differences in conditional value functions and conditional choice probabilities. For the case of logit errors, the inversion is derived by differencing equation (1.6) across j :

$$(1.8) \quad \ln \left(\frac{p_t(j, k)}{p_t(j', k)} \right) = \delta_t(j, k) - \delta_t(j', k).$$

Rewrite the Hotz-Miller inversion for the crop choice model as a relationship between ex ante current profits, continuation profits, and conditional choice probabilities:

$$(1.9) \quad \bar{\pi}_t(j, k) - \bar{\pi}_t(j', k) - \ln \left(\frac{p_t(j, k)}{p_t(j', k)} \right) = \beta E_t [\bar{V}_{t+1}(\kappa^+(j', k))] - \beta E_t [\bar{V}_{t+1}(\kappa^+(j, k))]$$

In binary choice logit models, the conditional choice probability term has a very simple interpretation: $\ln \left(\frac{p_t(j, k)}{p_t(j', k)} \right)$ is equal to the cutoff $\Delta \nu_t^*$ such that if $\nu_{jit} - \nu_{j'it} \geq \Delta \nu_t^*$, field i will be in land use j , and otherwise field i will be in land use j' . Thus, the left-hand-side of equation (1.9) expresses the minimum difference in expected profits during period t which justifies the choice of land use j rather than j' in period t , given a particular field state k . The right hand side expresses the expected loss in continuation values resulting from the choice of j instead of j' .

The next step is to replace the expected difference in continuation values with its realization and expectational errors:

$$(1.10) \quad \begin{aligned} \bar{\pi}_t(j, k) - \bar{\pi}_t(j', k) - \ln \left(\frac{p_t(j, k)}{p_t(j', k)} \right) &= \beta (\bar{V}_{t+1}(\kappa^+(j', k)) - \bar{V}_{t+1}(\kappa^+(j, k))) \\ &\quad + \varepsilon_t^V(j', k) - \varepsilon_t^V(j, k) \end{aligned}$$

where

$$\varepsilon_t^V(j, k) \equiv \beta (E_t [\bar{V}_{t+1}(\kappa^+(j, k))] - \bar{V}_{t+1}(\kappa^+(j, k))).$$

At this point, it is worth noting the importance of the perfectly competitive setting. Given Assumption 1, the law of iterated expectations implies that $E[\varepsilon^V(j, k) Z_t | \omega_t] = 0$ for any variable Z_t in the time t information set. In a model with strategic interactions, the same statement would only be true conditional on actions taken. That is, one could write $E[\varepsilon^V(j_{it}, k_{it}) Z_t | \omega_t, j_{it}] = 0$, but it would be necessary to condition on the action taken (j_{it}), and we would not have well-defined expectational error terms at the market level, which is what allows equation (1.10) to be estimated directly.

The final step in constructing the regression equation amounts to replacing differences in continuation values ($\bar{V}_{t+1}$) with terms that will cancel. To do this, I use a convenient relationship between ex ante and conditional value functions, which can be derived by adding and subtracting $\delta_t(j, k)$ from equation (1.7), and substituting using equation (1.6):

$$(1.11) \quad \forall j : \bar{V}_t(k) = \delta_t(j, k) - \ln(p_t(j, k)) + \gamma.$$

Equation (1.11) is a special case of Lemma 1 in Arcidiacono and Miller (2011), and versions of it also appear in Altug and Miller (1998), Arcidiacono and Ellickson (2011), and Murphy (2012).

Note well that equation (1.11) holds for any land use j . It is particularly convenient to apply equation (1.11) with j set equal to a *renewal action* j_{re} , where j_{re} satisfies $\kappa^+(j_{re}, k) = \kappa^+(j_{re}, k')$ for any field states k and k' . Choosing a renewal action for two different fields in period $t + 1$ will bring the fields into the same field states in period $t + 2$, regardless of what field states they were in period $t + 1$.¹⁴

Replacing the continuation values in equation (1.10) using equation (1.11),

$$(1.12) \quad \begin{aligned} \bar{\pi}_t(j', k) - \bar{\pi}_t(j, k) - \ln \left(\frac{p_t(j, k)}{p_t(j', k)} \right) &= \beta (\delta_{t+1}(j_{re}, \kappa^+(j', k)) - \delta_{t+1}(j_{re}, \kappa^+(j, k))) \\ &\quad - \beta \left(\ln \left(\frac{p_{t+1}(j_{re}, \kappa^+(j', k))}{p_{t+1}(j_{re}, \kappa^+(j, k))} \right) \right) \\ &\quad + \varepsilon_t^V(j', k) - \varepsilon_t^V(j, k). \end{aligned}$$

¹⁴Renewal actions are a special case of *finite dependence*, defined by Arcidiacono and Miller (2011). See Arcidiacono and Ellickson (2011) for further discussion of how finite dependence leads to relatively simple estimators based on conditional choice probabilities.

The conditional value function terms for period $t + 1$ could be written as profits in period $t + 1$ plus continuation values in period $t + 2$. However, because j_{re} is a renewal action, the continuation values in period $t + 2$ cancel, leaving

$$(1.13) \quad \delta_{t+1}(j_{re}, \kappa^+(j', k)) - \delta_{t+1}(j_{re}, \kappa^+(j, k)) = \bar{\pi}_{t+1}(j_{re}, \kappa^+(j', k)) - \bar{\pi}_{t+1}(j_{re}, \kappa^+(j, k))$$

The Euler equation comes from substituting equation (1.13) into equation (1.12):

$$(1.14) \quad \begin{aligned} \bar{\pi}_t(j, k) - \bar{\pi}_t(j', k) - \ln\left(\frac{p_t(j, k)}{p_t(j', k)}\right) &= \beta(\bar{\pi}_{t+1}(j_{re}, \kappa^+(j', k)) - \bar{\pi}_{t+1}(j_{re}, \kappa^+(j, k))) \\ &\quad - \beta \ln\left(\frac{p_{t+1}(j_{re}, \kappa^+(j', k))}{p_{t+1}(j_{re}, \kappa^+(j, k))}\right) \\ &\quad + \varepsilon_t^V(j', k) - \varepsilon_t^V(j, k). \end{aligned}$$

As explained above, the left-hand side represents the difference in profits necessary to justify the choice of land use j over land use j' in period t . Now, the right hand-side represents the expected (discounted) difference in profits in period $t + 1$ when an action is taken which compensates for the impact of the period t land use on the field state.

An interesting feature of this construction is that we can relate a difference in profits at time t to the difference in profits at time $t + 1$ given a renewal action, even though the renewal action is not always optimal. This is possible thanks to equation (1.11), which allows one to forward calculate the unconditional value function at time $t + 1$ using any action at time $t + 1$. As shown in Appendix 1.8.6, it turns out that the Hotz-Miller inversion makes this possible generally (not just for the assumption of logit errors).

Letting $j = crops$, $j' = other$, and $j_{re} = crops$, the Euler equation (1.14) can be rearranged into the following linear regression equation:

$$(1.15) \quad Y_{tk} = \tilde{\Delta}\alpha_{0k} + \alpha_R \Delta R_t + \tilde{\Delta}\xi_{tk} + \Delta\varepsilon_{tk}^V$$

where

$$\begin{aligned} Y_{tk} &\equiv \ln\left(\frac{p_t(crops, k)}{p_t(other, k)}\right) + \beta \ln\left(\frac{p_{t+1}(crops, 0)}{p_{t+1}(crops, \kappa^+(other, k))}\right) \\ \tilde{\Delta}\alpha_{0k} &\equiv \alpha_0(crops, k) - \alpha_0(other, k) \\ &\quad + \beta(\alpha_0(crops, 0) - \alpha_0(crops, \kappa^+(other, k))) \\ \Delta R_t &\equiv R_{crops, t} - R_{other, t} \\ \tilde{\Delta}\xi_{tk} &\equiv \xi_{crops, k, t} - \xi_{other, k, t} + \beta(\xi_{crops, 0, t+1} - \xi_{crops, \kappa^+(other, k), t+1}), \\ \Delta\varepsilon_{tk}^V &\equiv \varepsilon_t^V(crops, k) - \varepsilon_t^V(other, k). \end{aligned}$$

Notice that the dependent variable $Y_{t,k}$ can be constructed from estimated conditional choice probabilities and the discount factor. This calls for a two-stage estimation procedure: first, estimating conditional choice

probabilities to construct an estimate of $Y_{t,k}$; then, estimating the linear regression equation above with the estimated dependent variable.

Some comments on identification are in order. First, exclusion restrictions on the composite error term $\tilde{\Delta}\xi_{tk} + \Delta\varepsilon_{tk}^V$ are needed for estimation. Because the expectational error term $\Delta\varepsilon_{tk}^V$ is mean-uncorrelated with any variables in the information set ω_t by construction, it satisfies standard exclusion restrictions by construction.¹⁵ This same point was famously made by Hall (1978) in the context of consumption-savings decisions.

In contrast, substantive assumptions must be made about the unobservable shock term ξ to justify an estimator. For example, assuming $E(\tilde{\Delta}\xi_{tk}|\Delta R_t(k)) = 0$, ordinary least squares will deliver consistent estimates. If the unobservable shocks ξ are potentially correlated with observable returns R_t but uncorrelated with some observed variable in ω_t , then linear instrumental variables estimators could be used.

Another issue is whether the original intercept terms $(\alpha_0(j,k))$ can be recovered from the intercepts of the regression equation $(\tilde{\Delta}\alpha_{0k})$. This requires some restrictions; in fact, dynamic discrete choice models are generally not fully identified without some restrictions (Magnac and Thesmar, 2002). The fact that the model is not identified in its full generality is easy to see in this setting, for estimating equation (1.15) can only deliver $|\mathbf{K}|$ values of $\tilde{\Delta}\alpha_{0k}$, but there are $2|\mathbf{K}| - 1$ values of $\alpha_0(j,k)$ (after normalizing). The following assumption effectively limits the number of distinct values of $\alpha_0(j,k)$.

ASSUMPTION 3. The payoffs to non-cropland do not depend on the field state.

Given Assumption 3, the model is fully identified, and parameters of the payoff function can be recovered from regression equation estimates. Since it is harmless to rescale the payoff function by adding a scalar, it is generally possible to normalize $\alpha_0(j,k) = 0$ for a single choice of (j,k) . Assumption 3 implies that $\alpha_0(\text{other},k)$ does not depend on k , so I normalize $\alpha_0(\text{other},k) = 0$ for all k . With this normalization, $\alpha_0(\text{crops},k)$ can be recovered from $\tilde{\Delta}\alpha_{0k}$ with a little algebra.¹⁶

¹⁵To see this formally, let x_t represent some instrumental variable, and notice that

$$\begin{aligned} E[x_t \varepsilon_t^V(j,k)] &= \beta E[x_t (E[\bar{V}_{t+1}(\kappa^+(j,k))|\omega_t] - \bar{V}_{t+1}(\kappa^+(j,k)))] \\ &= \beta E[(E[x_t \bar{V}_{t+1}(\kappa^+(j,k))|\omega_t] - x_t \bar{V}_{t+1}(\kappa^+(j,k)))] \\ &= 0 \end{aligned}$$

where the second equality follows as long as x_t is within the time t information set, and the final equality follows from the law of iterated expectations.

¹⁶Specifically, $\alpha_0(\text{crops},k)$ can be recovered from $\tilde{\Delta}\alpha_{0k}$ as follows:

$$\alpha_{00} = \beta^{\bar{k}} \tilde{\Delta}\alpha_{0\bar{k}} + \sum_{k=0}^{\bar{k}-1} \beta^k (1 - \beta) \tilde{\Delta}\alpha_{0k},$$

and the other $\alpha_0(\text{crops},k)$ parameters could be solved for by proceeding forward through k , e.g., $\alpha_0(\text{crops},1) = \beta^{-1}((1 + \beta)\alpha_0(\text{crops},0) - \tilde{\Delta}\alpha_0(0))$.

1.3. Data and measurement

In this section, I describe the two data inputs used to estimate a model of US cropland. First, I explain how I constructed a unique panel data set on land use in the United States, providing the choice data inputs. Additionally, I explain the several steps involved in constructing a measure of expected returns R , which was taken for granted in Section 1.2.

While the effects of biofuels mandates (as well as any other policies which affect agricultural markets) depend on supply elasticities around the world, the United States is a particularly important player in global agricultural markets, for the US is the world's top exporter of agricultural products in general and cereal crops in particular.¹⁷ By any measure, US crop production is a large industry, with \$143 billion in sales in 2007 (US Census of Agriculture). Among the world's four most important crops (wheat, rice, corn, and soybeans, which account for 75% of crop production worldwide in caloric terms), US output accounts for roughly 23% of worldwide output in caloric terms (Roberts and Schlenker, 2013).

1.3.1. Land use data. The National Agricultural Statistics Service's Cropland Data Layer (CDL) program features finely detailed land cover data for the United States.¹⁸ The CDL's land cover classifications include over 50 different crops as well as about 20 non-crop classifications (e.g., grassland, forest, water, several levels of developed land). Since 2008, the annual CDL data have provided field-level resolution (30m or 56m) of the entire United States.

I construct a panel of land use outcomes using the CDL data for 2006-2012. Fields are defined as points on an 840m sub-grid of the CDL's grid. The CDL covers the entire contiguous US 2008-2012, but only some states in 2006 and 2007. Consequently, my panel is unbalanced, with 5-7 land use observations per field. I avoid CDL data from before 2006 because they are less reliable and cover fewer states.

After excluding water, protected land, and developed land, remaining points were classified as cropland or non-cropland. Table 4 lists the share of points in cropland by point and the the initial year in the panel by state.

This outside option lumps together several alternative land uses, which is partly a simplification and partly due to limitations in the data. The CDL data performs poorly when it comes to distinguishing among

¹⁷Based on export values, FAOSTAT Database on Agriculture, visited 10/31/2012.

¹⁸See <http://nassgeodata.gmu.edu/CropScape/> for data and Boryan et al. (2011) for a description data inputs, classification system, and validations procedures.

unmanaged grassland, pasture, and hay.¹⁹ These land uses are lumped together into the “other” land use category, which also includes other forms of unmanaged land such as shrubland, forests, and wetlands.

Further details regarding the land use data are included in Appendix 1.9.1.

1.3.2. Expected returns. Before constructing a measure of the average returns to cropland, I construct expected returns separately for each of eleven crops: corn, sorghum, soybeans, winter wheat, durum wheat, other spring wheat, barley, oats, rice, upland cotton, and pima cotton (denote this set of crops by $\mathbf{C}$). These crops account for about 94% of harvested cropland (excluding hay) in the US, according to the 2007 Census of Agriculture.

As mentioned in Section 1.2, expected returns should be a reflection of farmers’ incentives during planting season, when land managers must commit to a particular land use. I bring together data on prices received by farmers, futures prices during planting season, costs, yields, and weather to construct expected returns at the county level.

In the models I estimate, each US county defines an observable type of field. In some specifications, I allow for unobservable within-county heterogeneity in fields’ parameters, but the returns variable is always measured at the county level. Letting z index counties, the expected returns to planting crop c in county z in year t are given by

$$(1.16) \quad R_{czt} = (P_{czt} - e_{czt}) \cdot YIELD_{czt}$$

where P_{czt} represents the expected output price (including government payments), $YIELD_{czt}$ is the expected yield (output per acre), and e_{czt} represents a unit production cost.²⁰

The unit cost variable e_{ctz} is obtained from cost and returns estimates published by the USDA Economic Research Service, computed by dividing average operating costs by average yield at the level of ERS Resource Regions. Since many costs are incurred relatively early in the planting year, it may be reasonable to use realized cost data in constructing expected returns.

The expected price variable P_{czt} is constructed using futures prices, historical state-level prices received, and information on government payments. County-level expected yields $YIELD_{czt}$ are based on historical yields and weather data. In the remainder of this section, I provide an overview of how expected prices and yields are constructed. Further details can be found in Appendix 1.9.

¹⁹The distinctions among grassland, hay, and pasture have more to do with management practices than the type of plant covering the ground, so it is unsurprising that these land use types are difficult to tell apart from space. Unfortunately, these difficulties go beyond satellite scan data. Hay and pasture are hard to deal with in general because “hay farming” encompasses a diverse set of farming practices and plant species. Several quite different grasses and legumes are all considered hay, and it’s not clear where to draw the line between hay and pasture – e.g., how should we classify a field of grass which is both harvested and used for grazing livestock?

²⁰Strictly speaking, equation (1.16) assumes more than just risk-neutrality, for it ignores the covariance between realized prices and yields. However, this covariance is plausibly small for an individual farmer.

1.3.3. Expected yields. Expected yields are based on historical county-level yield estimates from the National Agricultural Statistics Service (NASS) together with detailed historical weather data.²¹

I estimate a model of yields similar to those estimated by Schlenker and Roberts (2009):

$$(1.17) \quad \ln(YIELD_{czt}) = \theta_{cz} + \theta_{cw}W_{zt} + \theta_{ct}t + \varepsilon_{czt}^Y$$

where W_{zt} includes weather variables described in Appendix 1.9, θ_{cz} is a county-level fixed effect, and θ_{ct} is a linear time trend. I estimate the model using 1981-2005 data. Separate models are estimated for each crop and ERS production region.

County-level expected yields are simply fitted values to equation (1.17) using the county's historical mean values of W_{zt} .

For some county-crop pairs, NASS yield data was unavailable (or only available for a few years). For such county-crop pairs, I impute fixed effects (θ_{cz}) based on a weighted average of the fixed effects of nearby counties, as long as some other county within 160km were included in the fixed effects regression (see Appendix 1.9.3 for details).

Figures 1.4-1.6 illustrate that my expected yields variables are excellent forecasts of average county-level yields between 2006 and 2011 (noting that no data from 2006-2011 were used to make the forecasts). The yields forecasts prove effective even for counties with fixed effects imputed from neighbors.²²

1.3.4. Expected prices. Expected prices are based on futures contract prices for corn, soybeans, wheat, and oats, and cotton from the Chicago Board of Trade (CBOT) and New York Board of Trade (NYBOT) together with information about government payments.

I use a simple forecasting model to map futures prices on commodity exchanges to expected prices received by farmers. I estimate the following equation using 1997-2011 data:

$$(1.18) \quad P_{cst}^{rec} = \theta_{1cs} + \theta_{2cs}P_{ct}^{fut} + \varepsilon_t$$

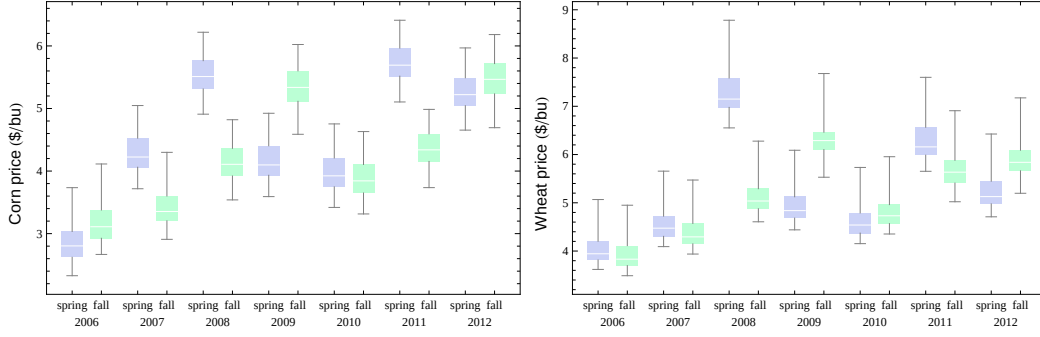
where P_{cst}^{rec} is the price received for crop c in state s in year t , P_{ct}^{fut} is a reference price.

The reference price P_{ct}^{fut} corresponds the price, during planting season, of a futures contract expiring in November or December of year t . The reference price contract for each crop is listed in Table 3.

²¹I thank Wolfram Schlenker and Michael Roberts for sharing this weather data.

²²For some county-crop pairs, NASS estimates feature little or no county-level information about yields prior to 2005, but some information about yields since 2006 – these are the crop-county pairs which I'm able to plot in Figures 1.4-1.6 although they were not included in the fixed effects regressions. For many county-crop pairs with imputed fixed effects, no validation data is available.

FIGURE 1.1. Expected market prices over time



Expected prices in spring are forecasts for November-December of the same calendar year. Expected prices in fall are for November-December of the following calendar year. Boxes span the median 50% of expected market prices by state. Whiskers span the full distribution of expected market prices within a year.

Expected market prices are defined as fitted values to equation (1.18):

$$(1.19) \quad P_{cst}^m \equiv \hat{\theta}_{1cs} + \hat{\theta}_{2cs} P_{ct}^{fut}.$$

Since the expected returns variable represents land owners' incentives at the time when land use decisions are made, it should coincide with planting season (or precede it slightly, since crop insurance sign ups, fertilizer application, and some input purchases precede the actual planting process). Unfortunately, there is considerable heterogeneity with respect to the timing of crop planting. The biggest difference comes from the fact that most crops are planted in the spring, but some are planted in the fall, including winter wheat. To deal with differences in planting seasons, I actually estimate two versions of equation (1.18), corresponding to the two different planting seasons. Expected market-level prices at the county level are taken from one version or the other, depending on which crops are typically planted in the county (see Appendix 1.9 for details).

Figure 1.1 displays the distribution (over US states) of corn and wheat expected market prices for each year and planting season.

Because of government payments, expected market prices do not fully represent the effective price per unit of output that a farmer can expect. I take into account counter-cyclical payments which kick in when

market prices fall below a target prices.²³ The expected price received by farmers is

$$P_{czt} \equiv \max \{P_{czt}^m, TP_{ct}\},$$

where TP_{ct} denotes the target price for crop c in year t as reported by the Farm Service Agency.

1.3.5. Aggregating over crops. To construct an aggregate measure expected returns to planting crops, I take a weighted average of expected returns:

$$(1.20) \quad R_{crops,z,t} = \frac{\sum_{c \in \mathbf{C}_z} A_{cs} R_{czt}}{\sum_{c \in \mathbf{C}_z} A_{cs}}$$

where s denotes the US state containing county z , A_{cs} is the average annual harvested area of crop c in state s (from 1981-2006), and $\mathbf{C}_z$ is the set of crops for which I am able to construct R_{czt} .

Differences in the set of crops used to construct expected returns $\mathbf{C}_z$ across counties is a potential cause for concern. This problem is alleviated to some extent by computing yields forecasts even for counties in which historical yield data for crop c is limited or non-existent (as described in Section 1.3.3). However, this strategy relies on imputing intercept terms based on intercept terms for nearby counties, and it would be unrealistic to extend the definition of nearby counties too far.

To ensure that the set $\mathbf{C}_z$ does not vary too much across counties within a given US state, a county z is included in my sample only if

$$(1.21) \quad \frac{\sum_{c \in \mathbf{C}_z} A_{cst}}{\sum_{c \in \mathbf{C}} A_{cst}} > .9$$

for every year t . In other words, I drop county z if I am not able to construct R_{czt} for crops which account for at least 90% of the acreage within $\mathbf{C}$ at the state level.

Other counties are excluded from my estimation because the crops in $\mathbf{C}$ constitute a small fraction of the crops planted in those counties (see Appendix 1.9.5 for details).

After all exclusions, there are 1990 counties remaining in my sample; collectively, they contain over 90% of US cropland, according to the 2007 Census of Agriculture. Table 4 displays the number of counties in the sample by US state, and Figure 1.2 is a map illustrating which counties are in the sample.

²³In a previous version of the paper, direct payment rates (government subsidies which do not depend on market conditions) were also included, but in most cases such payments do not depend on the decision of whether to plant crops or not in the US. If a farmer is eligible for direct payments, she will generally receive them even if she doesn't actually plant crops (as long as her land is still used for agricultural purposes). Similarly, planting new acres will not immediately increase the direct payments a farmer is eligible to receive.

1.4. Estimation

The model presented in Section 1.2 was focused on a set of fields which shared profit function parameters α , observable returns R_t and unobservable shocks ξ_t . To extend the model to have observable field-level heterogeneity, I simply index these variables by the observable type (α_z , R_{zt} , and ξ_{zt}), noting that the regression equation can be constructed separately for each type z .

Consistent with the notation in Section 1.3, observable characteristics z correspond to counties. Observable characteristics could also be based on soil characteristics and weather patterns, but the variation in these variables within-county is relatively small.²⁴ In Section 1.4.2, I explain how I extend the model to observable heterogeneity.

1.4.1. Truncation and smoothing of choice probability estimates. Conditional choice probability estimates can, in principle, be obtained directly from choice data as frequencies estimates. However, a problem with frequency estimates is the possibility that some estimated frequencies will be zero or one, in which case the Hotz-Miller inversion is not well-defined (for logit errors or any other distributional assumption with full support), and the regression equation cannot be constructed as described above.

The simplest way to avoid conditional choice probabilities estimates of zero and one is to truncate them, or replace any frequencies of zero (one) with an estimate which is very close but slightly larger (less) than zero (one). Specifically, I use the following truncation formula:

$$(1.22) \quad \hat{p}_{zt}(crops, k) = \max \left\{ \min \left\{ \hat{p}_{z't}^{freq}(crops, k), \frac{1}{2(N+10)} \right\}, 1 - \frac{1}{2(N+10)} \right\}$$

where $\hat{p}_{zt}^{freq}(crops, k)$ represents the frequency estimate of conditional choice probability $p_{zt}(crops, k)$. Assuming that true values of conditional choice probabilities fall in the interior of the unit interval, the probability of having to truncate a CCP falls to zero as the number of fields grows, and so truncation does not affect consistency and other asymptotic properties of the estimation strategy.

Another possibility is to smooth choice probabilities across neighboring counties. Even when extremal frequency estimates are not an issue, smoothing conditional choice probabilities may reduce sampling variance.²⁵ However, smoothing has downsides in that it will mechanically introduce spatial autocorrelation and could bias downward the responsiveness of choice probabilities to variables which shift the returns to planting crops.

²⁴Initially, I experimented with using only agro-climactic characteristics to define field types, but I found that there were persistent differences in yields across counties which could not be explained by the underlying agro-climactic characteristics of the fields within counties. Thus, counties plausibly capture some important spatial economic characteristics.

²⁵In the context of estimating entry/exit games, Pakes et al. (2007) show that smoothing across state variables improves the finite sample performance of their estimators substantially.

In specifications without unobservable heterogeneity, my smoothing procedure amounts to taking a weighted average of frequency estimates, with weights inversely proportional to distances between counties. The smoothed estimates are:

$$(1.23) \quad \hat{p}_{zt}(crops, k) = \frac{\sum_{z' \in Z_s} w_{zz'} \hat{p}_{z't}^{freq}(crops, k)}{\sum_{z' \in Z_s} w_{zz'}}$$

where Z_s is the set of counties within US state s , and the smoothing weight $w_{zz'}$ is inversely proportional to the square of the distance between counties: $w_{zz'} \equiv (1 + d_{zz'}/4)^{-2}$, where $d_{zz'}$ is the distance between z and z' measured in kilometers. If z and z' are not within the US state, then $w_{zz'} = 0$, even if they are adjacent.²⁶ Smoothing weights drop off very quickly; the median county has 93.3% of the total weight on its own frequency estimate, and the lowest own-county weight is over 84%.

After smoothing or truncating, I use the resulting conditional choice probabilities to construct dependent variables Y_{zt} , as in the derivation of equation (1.15). Table 6 compares results for CCP smoothing to CCP truncation.

1.4.2. Extension to unobservable heterogeneity. Ignoring unobservable heterogeneity could bias dynamic estimation results. For example, the persistence of a particular land use could be rationalized by appealing to switching costs or unobservable heterogeneity. When both factors are present, failure to account for unobservable heterogeneity may overstate the switching costs. While it's difficult to predict how counterfactuals will be affected by unobservable heterogeneity *ex ante*, it's worth assessing how incorporating unobservable heterogeneity affects the results, if only to evaluate its importance in obtaining reliable estimates of agricultural supply responses.

As described above, the empirical approach without unobservable heterogeneity involves first estimating conditional choice probabilities, then constructing a regression equation. With unobservable heterogeneity, I follow the same two steps, but the first-stage estimation of CCP's involves the estimation of a mixture model using the Expectation-Maximization (EM) algorithm, which Arcidiacono and Miller (2011) introduced to the estimation of dynamic discrete choice models with unobservable heterogeneity.²⁷

I assume that each field has a persistent unobservable binary characteristic: $\zeta_i \in \{0, 1\}$. We can think of $\zeta_i = 1$ as indicating that field i is relatively well suited to cultivation. In contrast, fields with $\zeta_i = 0$ might have poor soil or steep terrain features which would make it difficult to cultivate the field.

²⁶I measure the distance between counties in terms of the distance between centroids of the counties. The centroid of a county was calculated by averaging the coordinates of all points in my sample.

²⁷Also see McLachlan and Peel (2000) for a textbook treatment of the EM algorithm and Dempster et al. (1977) for the seminal reference.

Some new parameters are involved in the estimation of the mixture model. Each county has an unrestricted joint distribution of unobservable types and field states in the initial period:

$$(1.24) \quad \mu_z(\zeta, k) \equiv \Pr(k_{i1} = k, \zeta_i = \zeta | i \in I_z)$$

where I_z is the set of fields in county z . Furthermore, conditional choice probabilities must now be indexed by the unobservable characteristics $(p_{z\zeta t}(j, k))$.

Identification of dynamic discrete choice models with unobservable heterogeneity is a relatively new topic, so it's worth commenting on how this model is identified. Conditional on a field's state in period t , lagged land use decisions should have no impact on its current land use decision. However, we might observe in the data that fields with a given state k at time t are more likely to be in crops if they have also been in crops for each of the five years before t . If we're confident in the model of field states, then the persistence of cropland beyond what can be explained by field states may be evidence of unobservable heterogeneity. Thus, identification intuitively comes from correlations in agents' decisions over time which are not explained by field states. Consequently, identification relies crucially on the model of field states.²⁸

The posterior distribution on the unobservable type ζ for a given field is a function of type-specific conditional choice probabilities, the initial distribution of unobservable types μ_z , and the field's land use history:

$$(1.25) \quad q_{i\zeta} \equiv \Pr(\zeta_i = \zeta | \mathbf{j}_i, \mathbf{k}_i) = \mu_z(\zeta, k_{i1}) \prod_{t=1}^T p_{z\zeta t}(j_{it}, k_{it})$$

where $\mathbf{j}_i = (j_{i1}, j_{i2}, \dots, j_{iT})$, and $\mathbf{k}_i = (k_{i1}, k_{i2}, \dots, k_{iT})$.

The EM algorithm iteratively updates estimates of μ , p , and q until convergence. Let the superscript (m) denote values at the m th iteration.

The E step updates posterior probabilities $\hat{q}_{i\zeta}^{(m)}$ based on CCP estimates $\hat{p}_{z\zeta t}^{(m)}(\text{crops}, k)$ and prior probabilities $\hat{\mu}_{z\zeta}^{(m)}(k)$:

$$(1.26) \quad q_{i\zeta}^{(m)} = \hat{\mu}_{z\zeta}^{(m)}(k_{i1}) \prod_{t=1}^T \hat{p}_{z\zeta t}^{(m)}(j_{it}, k_{it}).$$

The M step goes in the other direction, estimating conditional choice probabilities and initial type probabilities, taking the posterior probabilities $q^{(m)}$ for granted. Conditional choice probabilities estimates

²⁸For my model, Kasahara and Shimotsu (2009), Proposition 4, implies that at least three periods of aggregate county-level CCP's (p_{zt}) are generically sufficient to identify type-specific CCP's $(p_{z\zeta t})$ and the initial type-state distribution function (μ_z) .

(at the m th iteration) can be computed as follows:

$$(1.27) \quad \hat{p}_{z\zeta t}^{(freq,m)}(j, k) = \frac{\sum_{i \in I_z} q_{i\zeta}^{(m-1)} D_{it}(j, k)}{\sum_{i \in I_z} q_{i\zeta}^{(m-1)} D_{it}(k)},$$

$$(1.28) \quad \hat{p}_{z\zeta t}^{(m)}(j, k) = \frac{\sum_{z' \in Z_s} w_{zz'} \hat{p}_{z'\zeta t}^{(freq,m)}(j, k)}{\sum_{z' \in Z_s} w_{zz'}},$$

where $D_{it}(j, k) = 1$ if $j_{it} = j$ and $k_{it} = k$ and $D_{it}(j, k) = 0$ otherwise; similarly for $D_{it}(k)$. Notice that $\hat{p}_{z\zeta t}^{(freq,m)}(crops, k)$ is the analog of a frequency CCP estimate, but weighted by posterior type probabilities. Naturally, $\hat{p}_{z\zeta t}^{(m)}(other, k) = 1 - \hat{p}_{z\zeta t}^{(m)}(crops, k)$.

I update the prior distribution of field states and unobservable types as follows:

$$(1.29) \quad \hat{\mu}_{z\zeta}^{(m)}(k) = \frac{\sum_{i \in I_z} q_{i\zeta}^{(m-1)} D_{i1}(k)}{\sum_{i \in I_z} q_{i\zeta}^{(m-1)}}.$$

The EM algorithm involves iterating on the E step (equation (1.26)) and the M step (equations (1.27-1.29)) until convergence.²⁹

The log-likelihood function for this mixture model can be written as follows:

$$(1.30) \quad \sum_z \sum_{i \in I_z} \log \left(\sum_{\zeta} \mu_z(\zeta, k_{i1}) \prod_{t=1}^T p_{z\zeta t}(j_{it}, k_{it}) \right).$$

In the special case where only own-county weights w_{zz} are non-zero, the above algorithm is indeed a traditional implementation of the EM algorithm. However, in general the algorithm is non-standard in that it does not converge to a local maximum of the likelihood function. In particular, the way conditional choice probabilities are updated in equations (1.27-1.28) does not necessarily increase the likelihood function, so iterations do not necessarily monotonically increase the likelihood function.

The lack of the EM algorithm's traditional monotonicity property removes the theoretical guarantee that the above algorithm will converge. However, as pointed out by Arcidiacono and Jones (2003), if it does converge (which it always does, in my experience), then it converges to values which satisfy equations (1.26-1.29). Then, these equations can be seen as defining a method-of-moments estimator, and the EM algorithm can be seen as a tool for implementing that method-of-moments estimator.

When using CCP truncation rather than smoothing, cross-county weights are set to zero so that standard frequency estimates are used in the EM algorithm. Consequently, the traditional monotonicity property holds. The truncation described by equation (1.22) is applied only after the EM algorithm has converged.

²⁹The terminal condition of I use requires that the likelihood function not change by more than 10^{-7} in absolute value for ten successive iterations.

1.4.3. Identifying assumptions. I do not model $R_{other,z,t}$ because the data on the returns to competing land uses is limited (especially for pasture land, which is the main competing land use for cropland in much of the US). Thus, $\Delta R_{zt} = R_{crops,z,t}$ in the model I estimate, effectively setting $R_{other,z,t} = 0$ and allowing the unobservable supply shock term to absorb variation in returns to competing land uses. Hereafter, I simply write R_{zt} to denote $R_{crops,z,t}$.

The returns to competing land uses are plausibly correlated with variation in crop returns both in the cross section and over time. For example, if a farmer's outside option is grazing cattle, then the unobservable shock in my model will be correlated with beef prices. Beef and crop output prices are correlated because feed grains are an important input in beef production. Therefore, the expected returns to grazing cattle (and the unobservable shock term) is likely correlated with crop returns over time. Furthermore, endogeneity problems in the cross section can result from the fact that land which is productive in growing crops is typically also productive in growing forage.

For these reasons, it is important to control for correlations between observed and unobserved returns in both the cross section and over time. To deal with correlations between unobserved and observed returns in the cross section, I include county-level fixed effects. To deal with correlations between unobserved and observed returns over time, I assume only that period- t levels of observed returns R are uncorrelated with subsequent changes in unobserved returns ξ , allowing for correlation between the levels of observed and unobserved returns at any given time. In the remainder of this section, I lay out the regression equations and identifying assumptions explicitly.

To simplify notation, let $n = (z, \zeta, k)$. A linear regression equation can be written out for a given choice of n :

$$(1.31) \quad Y_{nt} = \tilde{\Delta}\alpha_{0n} + \alpha_{R\zeta}R_{nt} + \tilde{\Delta}\xi_{nt} + \Delta\varepsilon_{nt}^V.$$

Equation (1.31) seems to call for a standard fixed effects estimation strategy. However, this would implicitly require dubious identifying assumptions.

The rational expectations assumption implies the moment

$$(1.32) \quad \forall t : E[\Delta\varepsilon_{nt}^V R_{nt}] = 0.$$

However, fixed effects estimation requires a stronger assumption:

$$(1.33) \quad \forall t, t' : E[\Delta\varepsilon_{nt}^V R_{nt'}] = 0,$$

which is not implied by the model and indeed unlikely to be true when considered carefully. For example, condition (1.33) requires the expectational error term for period t to be uncorrelated with returns in period $(t + 1)$. Recalling that the expectational error term is the difference between the time- t expectation of the time- $(t + 1)$ value function and its realization, of course returns in period $(t + 1)$ are one of the most important determinants of the expectational error term for period t .

A similar identification problem is considered by Arellano and Bond (1991), and their solution applies. Specifically, rather than using the standard fixed effects estimator, a first differences strategy can be used to remove the fixed effects, and the earlier values of explanatory variables (or further lagged values) can be used as instruments.

Formally, after taking first differences of equation (1.31),

$$(1.34) \quad Y_{n,t+1} - Y_{nt} = \alpha_{R\zeta} (R_{n,t+1} - R_{nt}) + (\tilde{\Delta}\xi_{n,t+1} - \tilde{\Delta}\xi_{nt}) + (\Delta\varepsilon_{n,t+1}^V - \Delta\varepsilon_{nt}^V).$$

We can then consistently estimate $\alpha_{R\zeta}$ using equation (1.34) with R_{nt} as an instrument for $(R_{n,t+1} - R_{nt})$ given the identifying assumption:

$$E [R_{nt} (\tilde{\Delta}\xi_{n,t+1} - \tilde{\Delta}\xi_{nt})] = 0,$$

noting that $E [R_{nt} (\Delta\varepsilon_{n,t+1}^V - \Delta\varepsilon_{nt}^V)] = 0$ follows from the rational expectations assumption.

The instruments I use are the lagged returns variable, expected caloric yields, and a constant term.

1.4.4. Static and myopic models. A myopic model features agents who lack forward-looking behavior but still allows for state dependence (i.e., the profit function may depend on the field state). In other words, a myopic model is a special case of the model described above with $\beta = 0$.

The regression equation for a myopic model (without unobservable heterogeneity) can be written as follows:³⁰

$$(1.35) \quad \ln \left(\frac{p_{zt}(crops, k)}{p_{zt}(other, k)} \right) = \alpha_{0zk} + \alpha_R R_{zt} + \xi_{ztk}.$$

Equation (1.35) makes it clear why ignoring forward looking behavior might lead to biased parameter estimates. If $\beta > 0$, the dependent variable in the regression equation implied by the model is

$$(1.36) \quad Y_{tk} = \ln \left(\frac{p_{z,t}(crops, k)}{p_{z,t}(other, k)} \right) + \beta \ln \left(\frac{p_{z,t+1}(crops, 0)}{p_{z,t+1}(crops, \kappa^+(other, k))} \right).$$

Thus, the dependent variable in the myopic model is missing a dynamic correction term if $\beta > 0$.

³⁰I have omitted difference operators to simplify the notation here, but the $\alpha_{0z}(k)$ in equation (1.35) should still be regarded as a function of differences in parameters of the payoff function (the $\alpha_{0z}(j, k)$ terms).

Static models are more restrictive than myopic models in that they also rule out state dependence – formally, a static model is a special case of the model in which the set of field states is degenerate – i.e., $\bar{k} = 0$ and $\mathbf{K} = \{0\}$. The regression equation for a static model is similar to equation (1.35), but with no dependence on k :

$$(1.37) \quad \ln \left(\frac{p_{zt}(\text{crops})}{p_{zt}(\text{other})} \right) = \alpha_{0z} + \alpha_R R_{zt} + \xi_{zt}.$$

1.4.5. Defining elasticities. Elasticities for static models (without unobservable heterogeneity) are computed as follows:

$$(1.38) \quad \left(\sum_z A_{zt} \right)^{-1} \sum_z \left(\frac{\partial A_{zt}}{\partial R_{zt}} (R_{zt'} - R_{zt}) \frac{P_{zt}}{P_{zt'} - P_{zt}} \right),$$

where A_{zt} represents the area of cropland in county z during year t , and P_{zt} is a weighted average of crop prices.³¹

While there is a natural definition of elasticities in static models, dynamic models feature many elasticities one might be interested in. For understanding the impacts of a long-run shift in demand like the US biofuels mandate, it is crucial to understand how farmers respond to a long-run change in prices. For dynamic models (including myopic models), I compute an aggregate long-run acreage-price elasticity as follows:

$$(1.39) \quad \left(\sum_z \sum_{\zeta} A_{z\zeta}^* (R_{zt}) \right)^{-1} \left(\sum_z \sum_{\zeta} (A_{z\zeta}^* (R_{zt'}) - A_{z\zeta}^* (R_{zt})) \frac{P_{zt}}{P_{zt'} - P_{zt}} \right)$$

Where $A_{z\zeta}^* (R)$ is the steady-state acreage of fields in county z of type ζ , given expected crop returns fixed at R indefinitely. When solving $A_{z\zeta}^*$, I assume that the unobservable shocks are fixed at the average values of the estimated annual shocks.

I report arc elasticities of acreage changes with respect to a 10% increase in all crop output prices. Formally, $t = 2012$ in equations (1.38-1.39), and t' is a counterfactual period in which expected output prices are 10% higher than their 2012 levels.

Calorie-price elasticities are calculated in the same manner as acreage-price elasticities, but with caloric yields multiplying acreages. Yields are taken to be exogenous (and fixed at 2012 levels) so that calorie-price elasticities reflect only extensive responses and not intensive responses. Caloric yields are weighted averages of crop-specific yields, using the same weights used for returns (equation (1.20)).

³¹County-level prices are averaged across crops using the same weights as county-level returns – see equation (1.3.5).

TABLE 1. Long-Run Price Elasticities for Different Specifications

	No Unobs. Heterogeneity		Two Types Per County	
	Acreage	Calorie	Acreage	Calorie
Static model ($\bar{k} = 0$)	-0.0012 (0.0358)	-0.0013 (0.0359)	0.0600 (0.0460)	0.0624 (0.0471)
Myopic models ($\beta = 0$)				
$\bar{k} = 1$	0.0703 (0.1427)	0.0693 (0.1416)	0.0037 (0.0380)	0.0054 (0.0388)
$\bar{k} = 2$	0.0171 (0.1839)	0.0160 (0.1820)	-0.0067 (0.0507)	-0.0063 (0.0500)
Dynamic models ($\beta = .9$)				
$\bar{k} = 1$	0.6669 (0.5470)	0.6144 (0.5031)	0.5449 (0.1764)	0.5990 (0.1944)
$\bar{k} = 2$	0.4757 (0.5151)	0.4519 (0.4932)	0.4152 (0.1610)	0.4425 (0.1724)

CCP truncation was used in the first stage and first differences with instruments were used in the second. Standard errors in parentheses.

1.5. Results

Table 1 presents long-run elasticity estimates for different assumptions on unobservable heterogeneity and field-level dynamics. While static analysis suggests that crop supply is highly inelastic – long-run acreage-price and calorie-price elasticities are less than .06 – dynamic models generate long-run elasticities from on the order of .4-.5.

Dynamic and myopic specifications differ only in the imputed discount factor (.9 and 0, respectively), but this impacts long-run elasticity estimates dramatically. Myopic models generate long-run elasticity estimates which are insignificantly different from zero.³²

In dynamic specifications, I effectively lose $\bar{k}$ periods of choice probability data from the beginning of my choice data sample because $\bar{k}$ periods of choices must be observed to infer field states. Furthermore, with $\beta > 0$, I cannot construct the dependent variable for the final period of observed choice data (see equation (1.15)). The most flexible dynamic specification I estimate features $\bar{k} = 2$, so with choice data spanning 2006-2012, the effective sample period used in my most flexible dynamic specification 2008-2011. In Table 1, the samples are harmonized across specifications so that differences in result do not reflect differences in the effective samples.

³² Standard errors are computed allowing for spatial and temporal autocorrelation. See Section 1.8.1 for details.

Differences between elasticity estimates between these different specifications can largely be explained by considering differences in dependent variables. First, consider the difference between a static model and myopic model without unobservable heterogeneity. While the dependent variable for the myopic model features choice probabilities which are specific to a certain field state,

$$(1.40) \quad \ln \left(\frac{p_{zt}(k)}{1 - p_{zt}(k)} \right),$$

the dependent variable for a static model is aggregated over field states:

$$(1.41) \quad \ln \left(\frac{p_{zt}}{1 - p_{zt}} \right) = \ln \left(\frac{\sum_k \mu_{zt}(k) p_{zt}(k)}{1 - \sum_k \mu_{zt}(k) p_{zt}(k)} \right),$$

where $\mu_{zt}(k)$ is the proportion of fields in state k within county z during year t . This aggregation can mute the apparent responsiveness of cropland to acreage changes. For example, it might be the case that the transition rate from non-cropland to cropland increases dramatically at high price levels. However, if the share of land in crops is typically small, the aggregate share of land in crops will increase only slightly when crop prices are high. Thus, a static model could predict a small acreage-price elasticity and fail to capture the elevated transition rate. In reality, the elevated transition rate might lead to a very large acreage response when prices are held at elevated levels for a long time. Thus, a static model effectively captures short-run correlations between acreage and returns in the data, and there's good reason to expect that these short-run correlations might understate actual long-run responses.

As noted in Section 1.4.4, dynamic and myopic models can be estimated from regression equations which are identical but for different dependent variables – the dependent variable in models with $\beta > 0$ includes an additional dynamic correction term. Furthermore, this correction term in dynamic models is a function of future conditional choice probabilities. Since agricultural commodity prices exhibit positive autocorrelation over time, current returns will generally be correlated with future returns, which are naturally correlated with future realizations of conditional choice probabilities. This means that the dynamic correction term cannot be written off as exogenous measurement error, and it should not be surprising that forward-looking dynamic specifications yield different parameter estimates than their myopic counterparts.

Table 2 provides a more detailed summary of estimation results for dynamic models. For models with unobservable heterogeneity, it is the low field types (those with a lower probability of being planted in crops) which have a higher value of the sensitivity parameter α_R . High field types tend to be in cropland most of the time, and the results reflect that their choice probabilities respond little to variation in expected returns.

TABLE 2. Dynamic Model Estimates

State dependence $\bar{k}$	1	1	2	2		
Field types	1	2	1	2		
Type		High	Low		High	Low
α_R	0.1405 (0.0632)	0.0359 (0.0595)	0.3822 (0.0740)	0.0553 (0.0593)	-0.0041 (0.0706)	0.2333 (0.0788)
Average intercepts (α_0)						
$k = 0$	-0.3268 (0.1677)	1.1118 (0.1579)	-2.6980 (0.1962)	-0.0546 (0.1632)	0.9259 (0.1944)	-2.1024 (0.2170)
$k = 1$	-4.8413 (0.1677)	-0.1744 (0.1579)	-5.2899 (0.1962)	-2.3996 (0.1632)	-1.3241 (0.1944)	-1.7699 (0.2170)
$k = 2$	— —	— —	— —	-5.7812 (0.1632)	-1.7508 (0.1944)	-5.0839 (0.2170)
Type's share of fields	1.0000	0.2702	0.7298	1.0000	0.2571	0.7429
Share in crops	0.2496	0.8624	0.0226	0.2516	0.8772	0.0350
Long run acreage-price elasticity	1.1353 (0.4714)	0.4232 (0.1160)		0.4757 (0.5151)	0.4152 (0.1610)	
Long run calorie-price elasticity	1.0379 (0.4261)	0.4525 (0.1270)		0.4519 (0.4932)	0.4425 (0.1724)	
Counties	1978	1978		1978	1978	
County-years	8281	8281		6303	6303	

CCP truncation was used in the first stage, first differences with instruments were used in the second, and $\beta = .9$. second. R measured in \$100/acre. Standard errors in parentheses.

1.6. Implications for the US biofuels mandate

To give some meaning to my elasticity estimates and as a preliminary assessment of their policy implications, I revisit Roberts and Schlenker's (2013, hereafter "RS") evaluation of the US biofuels mandate.

RS estimate a worldwide supply elasticity in the neighborhood of 0.1 using a static specification. While dynamics is the difference that I want to highlight in this comparison, another major difference in the results is that they estimate a supply elasticity for the entire world, whereas I estimate a supply elasticity for the US. Of course, the supply elasticity could differ between the US and the rest of the world, meaning that differences between my estimates and RS involves more than the difference between static and dynamic specifications. On the other hand, my static results suggest even lower elasticities than RS report, so the comparison of our elasticities might actually understate the difference resulting from dynamic modeling.

As RS explain, the US biofuels mandate is met mostly by corn ethanol, and the corn used to produced this ethanol corresponds to a staggering 5% of worldwide caloric production of major crops.³³ Furthermore, it is arguably realistic to claim that the US biofuels mandate corresponds to a long-run increase in demand. While US Renewable Fuel Standards are scheduled to require increasing amounts of biofuels other than corn ethanol (which arguably have a smaller impact the human food and animal feed supply), the absolute quantity which may come from corn ethanol is not scheduled to decrease (US EPA, 2011).

The effective long-run increase in worldwide crop demand caused by the mandate is plausibly less than 5% because a substantial by-product of corn ethanol production is distillers dried grain (DDG), which is now popular in animal feed blends in the US. Thus, some portion of the corn diverted to the biofuels market actually comes back to the food and feed market. However, the DDG by-products correspond to about one third of the dry weight of the original corn inputs to ethanol production. Although the nutritional equivalence of unprocessed corn and DDG may be questioned, a natural assumption (and the assumption adopted by RS) is that one third of the corn used in ethanol production is effectively returned to the food and feed supply.

Thus, let's assume that the US biofuels mandate corresponds to a 3.33% increase in long-run global crop demand. Assuming a demand elasticity for crop calories of -0.05 (within the range of what RS estimate), RS's crop supply elasticity of 0.1 implies a 22% increase in prices and a 2.2% increase in crop acreage (assuming an acreage-price elasticity equal to the quantity-price elasticity). In contrast, the supply elasticities from my most flexible model (an acreage-price elasticity of .42 and calorie-price elasticity of .44) imply a 6.8% increase in prices, and a 3% increase in crop acreage.³⁴

In other words, replacing RS's supply elasticity with mine implies a 35% larger land use response and 70% smaller price increase as a result of the US biofuels mandate.

Around the world, government supports for biofuels production present a particularly unfortunate dichotomy. While the reduced price effect makes the assessment of biofuels subsidies seem more positive – especially given the burden increased food prices place on less developed countries – the increased land use is likely to mean increased greenhouse gas emissions due the release of terrestrial carbon as land is cleared.

³³It should be noted that this 5% figure does not include the crops used to produce the biofuels mandated by many other countries around the world.

³⁴These calculations hold yields fixed, effectively ignoring the component of supply elasticities which would result from intensification. There is some evidence to suggest that yield-price responses may be small if not negligible (Berry and Schlenker, 2011), implying that acreage responses (the extensive response) would be the main source of crop supply responses. I calculate the price impact of the biofuels mandate following the same formula as RS:

$$\Delta p = \frac{\Delta q}{\mathcal{E}_S - \mathcal{E}_D}$$

where Δp is the relative change in price, $\Delta q = 0.033$ is the relative change in demand resulting from the US biofuels mandate, and $\mathcal{E}_S$ and $\mathcal{E}_D$ are the calorie-price elasticities of supply and demand, respectively. The acreage effect is then given by $\mathcal{E}_A \Delta p$, where $\mathcal{E}_A$ is the acreage-price elasticity.

Moreover, it should be noted that the relatively small price increase following from my elasticity estimate is based on plugging a long-run supply elasticity into a static equilibrium calculation. Even if the long-run price effect of the US biofuels mandate is as small as this 6.8% figure suggests, it is possible that the mandate had a much larger short run price effect, consistent with the spike in prices in 2008.

1.7. Conclusion

This paper's main contribution is to formulate a flexible empirical approach for analyzing land use based on a model of dynamically optimizing landowners. The method is easily implemented, for I derive a linear regression equation which can be used to estimate the model (a construction of potential use in other single agent dynamic settings such as dynamic demand estimation). My empirical approach accommodates unobservable market-level shocks as well as unobservable heterogeneity, unavoidable difficulties in modeling land use at a disaggregated level.

Furthermore, I estimate long-run crop acreage elasticities for the United States based on a new land use panel data set. Relative to my results with forward looking landowners and unobservable field-level heterogeneity, I find that static and myopic models understate long-run acreage elasticities, and specifications without unobservable heterogeneity overstate acreage elasticities. A preliminary comparison of results suggests that static models understate the long-run land use effects (and indirect environmental costs) of biofuels mandates and overstate their long-run effects on food prices.

1.8. Technical Appendix

1.8.1. Standard errors with spatial and temporal dependence . Standard errors are based on Conley's (1999) treatment of spatial autocorrelation in a Generalized Method of Moments (GMM) framework, noting that each regression in this paper falls within the GMM framework (whether a first-differences, fixed effects, or OLS specification, and whether or not instruments are used). Furthermore, Conley's (1999) approach to calculating standard errors for data with spatial dependence is an extension of Newey and West's (1987) approach to dealing with temporal dependence, and it can be applied to deal simultaneously with both spatial and temporal dependence.

Conley (1999) and Newey and West (1987) develop asymptotic theory for covariance matrix estimators which use weights between "nearby" observations which decline as the distance (and/or time duration) between the observations decreases. Furthermore, the weights change as the sample size increases. In practice, it is more common to simply use unit weights with some cutoff determining what count as "nearby" observations (Conley, 2008).

Let $\hat{g}_{zkt}$ represent the fitted value of moments for a given observation for a particular regression. Because regressions are estimated separately for each type, I omit the subscripts for unobservable types ζ from the notation. To simplify the notation further, let $s = (z, k, t)$, let S represent the set of (z, k, t) included in the regression, and let N_s represent the number of elements in S . To give an explicit example, for the case of a first-differences regression with instruments (à la Arellano and Bond (1991)), fitted moments can be computed as follows:

$$\hat{g}_s = ((Y_{zk,t+1} - Y_{zkt}) - \hat{\alpha}_R (R_{z,t+1} - R_{zt})) \begin{pmatrix} 1 \\ R_{zt} \\ CYIELD_{zt} \end{pmatrix}$$

where $CYIELD_{zt}$ is the expected caloric yield for county z in period t .

In computing standard errors, I estimate the asymptotic covariance matrix as follows:

$$\hat{V} = N_s^{-1} \sum_{s_1 \in S} \sum_{s_2 \in S} K(s, s') \hat{g}_{s_1} \hat{g}_{s_2}'$$

where K is a uniform kernel which equals one when observations s and s' fall within the same period and have nearby locations, or when s and s' refer to the same (z, k) and fall in adjacent periods. Formally,

$$K(s_1, s_2) = \begin{cases} 1 & \text{if } z_1 = z_2, \text{ and } |t_1 - t_2| \leq 1 \\ 1 & \text{if } d(z_1, z_2) \leq 100 \text{ and } t_1 = t_2 \\ 1 & \text{if } st(z_1) = st(z_2) \text{ and } t_1 = t_2 \\ 0 & \text{otherwise} \end{cases}$$

where $d(z_1, z_2)$ represents the distance between the centroids of counties z_1 and z_2 in kilometers, and $st(z)$ represents the US state that county z falls within. Thus, two counties are counted as “nearby” if they are within 100 kilometers of each other or if they are in the same state.

Table 6 includes standard errors for estimates computed with and without corrections for autocorrelation. Standard errors on elasticities and estimated parameter values in all other tables allow for autocorrelation.

1.8.2. Standard errors on long run elasticities. Standard errors on long-run elasticity estimates are calculated by simulating the estimated asymptotic distribution of parameters. Formally, let $LRE(\alpha)$ represent the long run elasticity given the vector of parameters α . Means and standard errors for $LRE(\alpha)$ are calculated as follows:

$$\mu_{LRE} = \frac{1}{400} \sum_{n=1}^{400} LRE(\alpha_n)$$

$$SE_{LRE} = \sqrt{\frac{1}{399} \sum_{n=1}^{400} (LRE(\alpha_n) - \mu_{LRE})^2}$$

where α_n represent random draws from the estimated asymptotic distribution of α .

1.8.3. Recovery of profit function parameters. Before parameters of landowners' profits functions can be computed, estimates of fixed effects must be computed, and then the profit function parameters can be recovered as described in Section 1.2.2.

Estimates of the fixed effects are computed as follows:

$$\tilde{\Delta}\hat{\alpha}_{0z\zeta kt} = T_{z\zeta k}^{-1} \sum_t \hat{e}_{z\zeta kt}$$

where $\hat{e}_{z\zeta kt} \equiv Y_{z\zeta kt} - \alpha_{\zeta R} R_{z\zeta kt}$, and $T_{z\zeta k}$ is the number of observations of (z, ζ, k) .

Then, with the estimated regression equation intercepts $\tilde{\Delta}\hat{\alpha}_{0z\zeta kt}$ in hand, intercepts of the profit function $\hat{\alpha}_{0z\zeta kt}$ can be recovered as described in footnote 16.

1.8.4. Regression weights . Conditional choice probabilities will be more precisely estimated for field states with large numbers of fields in that state. Given estimates of choice probabilities and the distribution of field types from the first stage, I construct weights to be used in the second stage which are inversely proportional to the estimated standard error of the constructed dependent variable.

The same weights are use for a given (z, ζ, k) across periods t , calculated as follows:

$$\underline{p}_{z,\zeta,k} \equiv \frac{1}{T_z \sum_t \min \{ \hat{p}_{z,\zeta,t}(crops, k), \hat{p}_{z,\zeta,t}(other, k), \hat{p}_{z,\zeta,t+1}(crops, 0), \hat{p}_{z,\zeta,t+1}(crops, \kappa^+(other, k)) \}}$$

where

$$w_{z,\zeta,k}^Y \equiv \hat{N}_{z,\zeta,k} \underline{p}_{z,\zeta,k} / (1 - \underline{p}_{z,\zeta,k})$$

where T_z is the number of periods of CCP estimates for county z , and $\hat{N}_{z,\zeta,k}$ is the estimated average number of fields in field state k in county z of type ζ .

Table 6 illustrates the impact of incorporating regression weights on long run elasticities. Estimates in all other tables are based on regressions with weights.

1.8.5. First stage choice probability estimation. For specifications with unobservable heterogeneity, choice probabilities are estimated by alternating between the expectation step (equation (1.26)) and

maximization step (equations (1.27-1.29)) until the change in the likelihood function between full iterations is less than 10^{-7} for ten successive iterations. This first-stage estimation is run separately for each specification and US state.

1.8.6. Generalizing the regression equation derivation. In this section, I discuss the generality of the derivation of the regression equation (1.15). In particular, I consider three aspects of the model presented in Section 1.2: the binary choice setting, the renewal action, and logit errors (Assumption 2).

The choice set. The binary choice setting played no role in the derivation of the regression equation. The actions j and j' used throughout the derivation could be any two actions in a discrete choice set $\mathbf{J}$ of arbitrary size. In a multinomial setting, the main thing that changes is that there are more choices of (j, j') , each implying a different version of equation (1.15) (whereas it was without loss to set $j = \text{crops}$ and $j' = \text{other}$ in the binary choice setting).

Finite dependence. The requirement that a renewal action exists can also be relaxed substantially to the requirement that the evolution of field states satisfies finite dependence. The existence of a renewal action implies one-period dependence, where only one period is needed to harmonize the field states of two fields. Naturally, s -period dependence means that states can be harmonized within s periods. As discussed by Arcidiacono and Ellickson (2011) and Arcidiacono and Miller (2011), s -period dependence generally yields estimators which can be constructed with s -periods ahead CCPs. To extend my regression equation construction to s -period dependence, equation (1.11) must be applied iteratively s times – i.e., until the field states k are harmonized and continuation values cancel.

It should be noted that my derivation of a regression equation requires only finite dependence with respect to the field states k – market level state variables need not satisfy any such condition.

Idiosyncratic error terms. The distributional assumption on the idiosyncratic error terms (ν) can be relaxed, too. Specifically, Assumption 2 was used in two places: in the Hotz-Miller inversion (equation (1.8)), and in equation (1.11), which relates the ex ante value function ($\bar{V}(k)$) to an additively separable function of a conditional value function for a particular action ($\delta(j, k)$) and conditional choice probabilities. As shown below, the existence of an additively separable equation analogous to equation (1.11) is a general consequence of the Hotz-Miller inversion.

For completeness, I restate the Hotz-Miller inversion in my notation (see Hotz and Miller (1993), Proposition 1, for the original result). Here, I drop the assumption that the idiosyncratic error terms ν have a type 1 extreme value assumption, but maintain the assumption that ν has a known distribution F conditional on state variables. Conditional choice probabilities can be written as if a landowner faces a static random

utility model with mean utility from option j equal to the conditional value function $\delta_t(j, k)$:

$$(1.42) \quad \forall j \in \mathbf{J} : p_t(j, k) = \Pr(\forall j' \in \mathbf{J} : \delta_t(j, k) + \nu_{jt} \geq \delta_t(j', k) + \nu_{j't}).$$

Normalizing $\delta_t(j, k)$ to zero for some j , equation (1.42) defines a mapping $\phi : \mathbb{R}^{|\mathbf{J}|-1} \rightarrow \Delta^{|\mathbf{J}|-1}$ from (differences in) conditional value functions to conditional choice probabilities.

RESULT 1. (*The Hotz-Miller inversion*) Assuming F has a well-defined density function, ϕ is invertible.³⁵

One value of $\delta_t(j, k)$ must be normalized to zero for the mapping to be invertible, so the inversion effectively recovers differences in conditional values. We can define the Hotz-Miller inversion in terms of an arbitrary reference action $J \in \mathbf{J}$, and write $\phi_j^{-1}(\mathbf{p}_t(k)) = \delta_t(j, k) - \delta_t(J, k)$, where $\mathbf{p}_t(k) = \{p_t(j, k)\}_{j \in \mathbf{J}}$.

Proposition 2 shows that that equation (1.11) is a general consequence of the Hotz-Miller inversion. Therefore, the derivation of a linear regression equation can be generalized to different distributional assumptions on the idiosyncratic shocks. While the result is equivalent to Lemma 1 in Arcidiacono and Miller (2011), I offer a simpler proof.³⁶

RESULT 2. (*Arcidiacono and Miller, Lemma 1*) Assume F has a well-defined density function. For any $j \in \mathbf{J}$, there exists a function ψ_j such that

$$\bar{V}_t(k) = \psi_j(\mathbf{p}_t(k)) + \delta_t(j, k).$$

PROOF. Define

$$(1.43) \quad S(\delta) \equiv \int \max\{\delta_1 + \nu_1, \delta_2 + \nu_2, \dots, \delta_J + \nu_J\} dF(\nu).$$

For any real number c ,

$$\max\{\delta_1 + \nu_1 - c, \delta_2 + \nu_2 - c, \dots, \delta_J + \nu_J - c\} = \max\{\delta_1 + \nu_1, \delta_2 + \nu_2, \dots, \delta_J + \nu_J\} - c.$$

It follows that

$$S(\delta) = S(\delta_1 - \delta_J, \delta_2 - \delta_J, \dots, 0) + \delta_J.$$

³⁵Being more careful, the inverse function ϕ^{-1} is defined on the entire simplex Δ^{J-1} if the support of F is bounded, and defined only on the interior of Δ^{J-1} if F has full support, as is the case with the logit errors above (Norets and Takahashi, 2012).

³⁶I thank Steve Berry for suggesting this proof strategy.

Notice that the ex ante value function is given by $\bar{V}_t(k) = S(\delta_t(k))$ where $\delta_t(k) = (\delta_t(1, k), \dots, \delta_t(J, k))$. Thus,

$$(1.44) \quad \bar{V}_t(k) = S(\delta_t(1, k) - \delta_t(J, k), \delta_t(2, k) - \delta_t(J, k), \dots, 0) + \delta_t(J, k).$$

Using Proposition 1, substitute $\phi_j^{-1}(p_t(k)) = \delta_t(j, k) - \delta_t(J, k)$ into equation (1.44):

$$\bar{V}_t(k) = S(\phi_1^{-1}(p_t(k)), \phi_2^{-1}(p_t(k)), \dots, 0) + \delta_t(J, k).$$

Noting that J denotes an arbitrary element of $\mathbf{J}$, defining $\psi_J(p) \equiv S(\phi^{-1}(p), \phi^{-1}(p), \dots, 0)$ completes the result. $\square$

Equation (1.11) is a particularly simple special case of Proposition 2, with $\psi_j(\mathbf{p}) = -\ln(p_j) + \gamma$. In general, $\psi_j(\mathbf{p})$ may be a more complicated function of conditional choice probabilities (for instance, involving probabilities for more than one alternative), but the additively separable aspect of equation (1.11) is preserved, maintaining the possibility of constructing a linear regression equation.

1.9. Data appendix

1.9.1. Land cover panel. The accuracy of the CDL data has improved steadily, and CDL data is now an input in official USDA acreage estimates. The data has become especially accurate in recent years for major crops.³⁷ However, the data are still of limited use in distinguishing between certain similar land cover types, especially grassland, pasture, and hay.

My sample constitutes an 840m sub-grid of the CDL data, a level of spacing chosen to strike a balance between having a comprehensive sample of fields and artificially increasing the sample size by sampling many points from individual fields. Furthermore, the 840m grid scale facilitates matching of points across years when the source data's grid spacing changed from 56m to 30m. While the grid coordinates do not always match up exactly after such changes, one can always find sub-grids spaced by 840m such that the centroids of the two sub-grids are within 1m of each other in each dimension.

Points which were classified as developed land, water, or protected land (in any year) were excluded from the data set. This means that non-cropland includes mostly grassland, forests, and shrub land.

Points were assigned to counties using 2010 county boundary spatial data files provided by the Census Bureau. Spatial data on protected land was obtained from the Global Agro-Ecological Zones database, and associated with points in the CropScape data using nearest neighbor interpolation. All coordinate conversions and spatial merges were done with ArcGIS 10.

³⁷CDL accuracy data is available at <http://www.nass.usda.gov/research/Cropland/sarsfaqs2.html>

1.9.2. Weather data. The weather data used in this paper was generously provided by Wolfram Schlenker and Michael Roberts. These data are described in more detail in the data appendix to Schlenker and Roberts (2009).

Weather variables W_{zt} include degree days above 10 degrees Celsius, degree days above 30 degrees Celsius, precipitation, and interactions. The precipitation variable is simply the total precipitation from March to August. Degree days above T_{min} degrees Celsius are defined as:

$$DD_{T_{min}} = \int_{t_0}^{t_1} \text{Max}(T_t, T_{min}) dt$$

where T_t is the temperature at time t , t_0 is the beginning of March 1, and t_1 is the end of August 31. The integral is approximated using a sinusoidal interpolation between the high and low temperature each day.

Interactions of the degree days variables with precipitation were computed by multiplying daily values of degree days by daily precipitation values, then summing over the March-August period.

The weather variables are computed for 872505 grid points. Afterward, I simply average over grid points to form the county-level weather variables which are used in the yield regressions.

1.9.3. Yield regressions and forecasts. Tables 8-12 present regression results for sorghum, barley, oats, rice, and upland cotton (estimated separately for each ERS region). Because Schlenker and Roberts (2009) estimate similar models of corn, soybeans, and wheat yields (with similar results), I do not report the coefficient estimates for those regressions.

The results for these other crops are qualitatively similar to Schlenker and Roberts's estimates for corn, soybeans and wheat. Degree days above 30C typically have a strong negative effect on yields whereas degree days above 10C typically have a weaker positive effect on yields (except for oats and barley, where degree days above 10C seem to have a mild negative effect on yields).

I include a county-crop (z, c) in the yield regressions if NASS reports yields for crop c in county z and a harvested acreage of at least 100 acres for at least five years between 1997 and 2005.

For county-crop pairs (z, c) not included in the yield regressions, I impute a fixed effect θ_{cz} if some counties within 160km of county z were included in the yield regression for crop c . Imputed fixed effects are computed as follows:

$$\hat{\theta}_{cz} = \frac{\sum_{z' \in \mathbf{Z}_c} w_{zz'}^Y \hat{\theta}_{z'}}{\sum_{z' \in \mathbf{Z}_c} w_{zz'}^Y}$$

where $\mathbf{Z}_c$ is the set of counties included in the yield regression for crop c , and

$$w_{zz'}^Y = \begin{cases} (1 + d_{zz'}/10)^{-2} & \text{if } d_{zz'} \leq 160 \\ 0 & \text{otherwise} \end{cases}$$

TABLE 3. Reference prices

Reference price	Crops
CBOT corn	corn, sorghum
CBOT soybeans	soybeans
CBOT wheat	winter wheat, durum wheat, other spring wheat, barley, rice
CBOT oats	oats
NYBOT cotton	upland cotton, pima cotton

where $d_{zz'}$ is the distance between counties z and z' in kilometers (as measured by their centroids). These weights $w_{zz'}^Y$ are similar to those used in smoothing conditional choice probabilities ($w_{zz'}$). However, in this case I allow non-zero weights between counties within different states, but not between counties which are more than 160km apart or in different ERS regions.

1.9.4. Expected prices. All reference prices were based on contracts for December delivery, except for soybeans, which was based on November delivery contracts.

To deal with differences in planting seasons, I actually estimate two versions of equation (1.18), corresponding to the two different planting season. First, I estimate the model when P_{ct}^{fut} is the average closing price of the reference contract during February-March, capturing the relationship between received prices and futures contract prices before spring planting season (i.e., when most crops are planted). In the second version, P_{ct}^{fut} corresponds to closing prices during August-September of the previous year, capturing the relationship between received prices and futures contract prices before fall planting season (i.e., when winter wheat is planted).

Thus, I actually calculate two versions of state-level expected market prices, denoted by $P_{cst}^{m,sp}$ and $P_{cst}^{m,fa}$. I then assign these values to counties based on the composition of cropland as reported by USDA-NASS. Specifically, if at least 10% of a county's cropland is in winter wheat and rye every year between 2006-2011, then a county is designated as a fall planting county, and expected prices are based on futures prices in August-September of the previous year (i.e., $P_{czt}^m = P_{cst}^{m,fa}$). For other counties, expected prices are based on futures prices in February-March (i.e., $P_{czt}^m = P_{cst}^{m,sp}$).³⁸

Table 5 presents results using alternative measures of expected prices in which price forecasts were based on fall futures prices for all counties.

1.9.5. Counties in the sample. The county-specific set of crops $\mathbf{C}_z$ refers to the set of crops such that I am able to compute the expected yield for county z . I am able to construct the expected yield for

³⁸The threshold is chosen to err on the side of classifying counties as fall planting counties if substantial amounts of both spring and fall crops are planted in them. If winter wheat is an option farmers consider, they must at least make the decision about whether to plant winter wheat or not during the fall.

crop c in county z if (c, z) is included in the yield regression (as described in Section 1.9.3 above) or if there is another county z' within 160km of county c such that (c, z') is included in the yield regression.

A county is included in my sample only if three conditions hold:

- (1) Within my field-level panel, the crops in $\mathbf{C}$ comprise at least 25% of the county's total cropland every year.
- (2) Within my field-level panel, the county has at least 10 points cropland every year.
- (3) I am able to calculate $YIELD_{czt}$ for the prominent crops within county z 's state – specifically,
$$\frac{\sum_{c \in \mathbf{C}_z} A_{cst}}{\sum_{c \in \mathbf{C}} A_{cst}} > 9.$$

Furthermore, thirteen states were excluded from the analysis entirely either because they contain little cropland or have a relatively small share of cropland in the crops I model (Arizona, Connecticut, Delaware, Florida, Maine, Maryland, New Hampshire, Nevada, New Mexico, Rhode Island, Vermont, Virginia, West Virginia). No CDL data is available for Alaska and Hawaii.

According to the 2007 Census of Agriculture, the counties remaining in my sample account for over 90% of US cropland. Figure 1.2 presents a map of these counties.

1.10. Sensitivity analysis

1.10.1. The expected returns measure. The crucial explanatory variable in my second-stage estimation is the measure of expected returns. The construction of the measure of expected returns involves at least two substantive assumptions. First, expected prices were forecasted using futures prices around the beginning of planting season as explained in Section 1.9.4. However, farmers may effectively commit to planting crops substantially before this time. Anecdotally, fertilizer is often purchased several months before planting season, and preparing previously uncultivated land for planting certainly could take several months or more, depending on the condition of the terrain.

FIGURE 1.2. Map of sample counties

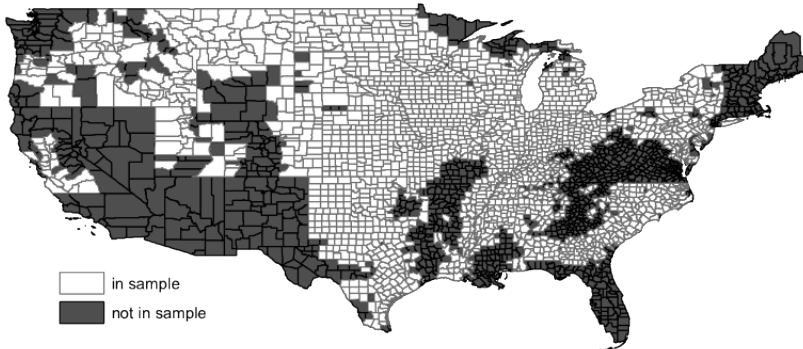


TABLE 4

SUMMARY STATISTICS FOR LAND USE PANEL				
State	counties in sample	points in sample	share in crops	initial year
Alabama	51	124896	0.050	2008
Arkansas	27	61160	0.515	2006
California	19	96977	0.227	2007
Colorado	30	187410	0.135	2008
Georgia	68	89023	0.170	2008
Idaho	38	203234	0.084	2007
Illinois	102	169199	0.726	2006
Indiana	92	109041	0.570	2006
Iowa	99	180953	0.706	2006
Kansas	105	277594	0.386	2006
Kentucky	60	68255	0.194	2008
Louisiana	32	71198	0.208	2006
Maryland	22	25390	0.237	2008
Michigan	65	128556	0.253	2007
Minnesota	78	201509	0.494	2006
Mississippi	49	89967	0.222	2006
Missouri	81	153122	0.301	2006
Montana	49	444292	0.078	2007
Nebraska	90	238056	0.358	2006
New Jersey	13	12405	0.111	2008
New York	49	138799	0.080	2008
North Carolina	77	126313	0.160	2008
North Dakota	53	225516	0.408	2006
Ohio	72	103091	0.460	2006
Oklahoma	60	187097	0.216	2006
Oregon	20	182161	0.046	2007
Pennsylvania	56	119015	0.107	2008
South Carolina	34	76609	0.094	2008
South Dakota	62	237054	0.317	2006
Tennessee	58	84205	0.157	2008
Texas	175	555467	0.178	2008
Utah	14	161940	0.012	2008
Washington	14	74547	0.255	2006
Wisconsin	65	156630	0.223	2006
Wyoming	11	113941	0.022	2008

As a preliminary robustness check on the assumptions about the timing of farmer's decisions, I also construct expected prices which are based on futures prices in the previous fall. For counties classified as fall planting counties, this does not change the measure of expected returns. This alternative measure of expected prices also removes much of the little cross-sectional price variation in the original measure – when price forecasts for different places are based on futures prices for different months, month-to-month variation in futures prices effectively creates some cross-sectional price variation.

A second assumption is that operating costs are linearly proportional to expected yields. An alternative assumption is that the costs per acre are fixed for a given crop within each of the ERS regions. These

TABLE 5. Long-run Elasticities for Different Measures of Returns

Acreage elasticity	0.4152 (0.1610)	0.4668 (0.1642)	0.3224 (0.1664)	0.3026 (0.1579)
Caloric Elasticity	0.4425 (0.1724)	0.4983 (0.1759)	0.3420 (0.1757)	0.3205 (0.1669)
Price forecasts use	Planting season futures		Futures from prev. fall	
Costs per acre	Proportional to yields	Flat within region	Proportional to yields	Flat within region

Long-run elasticities for models with two unobservable types and two periods of state dependence. CCP truncation was used in the first stage, first differences with instruments were used in the second, and $\beta = .9$. Standard errors in parentheses.

assumptions arguably represent extremes which are likely to bound the true proportionality between costs and yields. Plant nutrient requirements generally increase with the amount of plant growth, so it is plausible that costs should increase with yields.

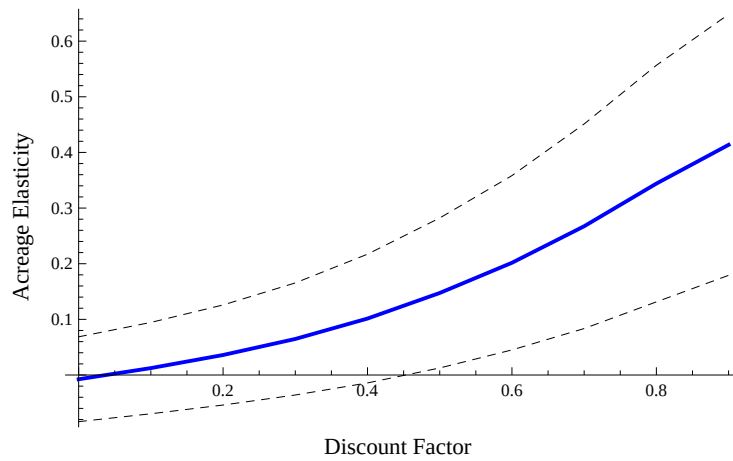
Table 5 shows how these two assumptions behind the construction of returns impact long-run elasticity estimates. Assumptions on the timing of the decision make a more substantial difference, increasing point estimates of long-run elasticities by 30-50%.

1.10.2. The discount factor. The discount factor for all dynamic models discussed above either use $\beta = .9$ (for models I call "dynamic") or $\beta = 0$ (for models I call "myopic"). As discussed by Rust (1987), discount factors in dynamic discrete choice models are often poorly identified, and I follow common practice in imputing a discount factor.

It is straightforward to estimated the model for different discount factor imputations, and as Figure 1.3 illustrates, I find that the relationship between long-run elasticity estimates and the discount factor tends to be increasing and convex.

1.10.3. Different estimation approaches. The most striking aspect of Table 6 is impact of autocorrelation on estimated standard errors, reflecting considerable spatial autocorrelation across counties. Spatial autocorrelation increases standard errors by more for specifications with smoothing than with truncation – this makes sense given that smoothing mechanically introduces spatial autocorrelation into the CCP estimates.

FIGURE 1.3. Long-Run Elasticities and the Discount Factor



Elasticities for model with two unobservable types and two periods of state dependence. CCP truncation was used in the first stage, and first differences with instruments were used in the second. Dashes indicate 95% confidence interval.

Elasticity estimates vary modestly based on whether fixed effects, first differences, or first differences with instrumental variables are used. Thus, the endogeneity problem discussed in Section 1.4.3 may be a relatively small issue in practice.

Using weights (as described in Section 1.8.4) leads to substantially larger elasticity estimates (although not statistically significantly larger). The difference between CCP smoothing and CCP truncation is somewhat more modest.

TABLE 6. Long-Run Acreage Elasticities for Different Estimation Approaches

CCPs	Weighted	Regression Approach		
		FE	FD	FDIV
smoothed	no	0.2909	0.2840	0.2540
		(0.0246)	(0.0211)	(0.0301)
		<i>(0.2376)</i>	<i>(0.1976)</i>	<i>(0.2506)</i>
smoothed	yes	0.4485	0.3803	0.4325
		(0.0236)	(0.0203)	(0.0278)
		<i>(0.2021)</i>	<i>(0.1632)</i>	<i>(0.2068)</i>
truncated	no	0.3657	0.3674	0.3068
		(0.0374)	(0.0318)	(0.0439)
		<i>(0.1972)</i>	<i>(0.1650)</i>	<i>(0.2253)</i>
truncated	yes	0.4136	0.3860	0.4152
		(0.0242)	(0.0211)	(0.0302)
		<i>(0.1519)</i>	<i>(0.1185)</i>	<i>(0.1610)</i>

Regression approaches are fixed effects, first differences, and first differences with instruments. All models feature two unobservable types, two periods of state dependence, and $\beta = .9$. Standard errors in parentheses standard errors with autocorrelation in italics.

TABLE 7

LAND COVER CLASSIFICATIONS					
CDL Classification	Mine	land cover %	CDL Classification	Mine	land cover %
Grassland/Herbaceous	other	18.03	Barley	crops	0.14
Shrub/Scrub	other	11.98	Sunflower	crops	0.11
Deciduous Forest	other	11.81	Dry Beans	crops	0.1
Evergreen Forest	other	8.31	Sugarbeets	crops	0.09
Corn	crops	8.04	Oats	crops	0.09
Soybeans	crops	6.26	Durum Wheat	crops	0.09
Pasture/Hay	other	5.24	Canola	crops	0.09
Developed, Open Space	excluded	3.83	Peanuts	crops	0.08
Woody Wetlands	other	3.6	Potatoes	crops	0.07
Winter Wheat	crops	3.05	Sod/Grass Seed	other	0.06
Pasture/Grass	other	2.82	Almonds	crops	0.05
Fallow/Idle Cropland	other	2.04	Peas	crops	0.04
Open Water	excluded	1.77	Grapes	crops	0.04
Non-alfalfa Hay	other	1.67	Millet	crops	0.04
Developed, Low Intensity	excluded	1.53	Rye	crops	0.04
Alfalfa	other	1.33	Lentils	crops	0.03
Cotton	crops	1.31	Walnuts	crops	0.03
Herbaceous Wetlands	other	1.29	Apples	crops	0.03
Spring Wheat	crops	1.18	Pecans	crops	0.02
Mixed Forest	other	0.78	Dbl. Crop WinWht/Sorghum	crops	0.02
Barren Land	other	0.74	Dbl. Crop WinWht/Cotton	crops	0.02
Developed, Medium Intensity	excluded	0.56	Sweet Corn	crops	0.02
Sorghum	crops	0.44	Aquaculture	excluded	0.02
Dbl. Crop WinWht/Soy	crops	0.41	Sugarcane	crops	0.02
Rice	crops	0.24	Clover/Wildflowers	other	0.02
Developed, High Intensity	excluded	0.2			
Percentages are for counties in my sample in 2011. Only land cover classifications with at least 1000 sample observations are listed above.					

TABLE 8. Sorghum yield regressions

VARIABLES	(1) region 1	(2) region 3	(3) region 4	(4) region 5	(5) region 6	(6) region 7	(7) region 8	(8) region 9
DD10	0.00041*** (0.000034)	0.0015*** (0.00013)	0.00038*** (0.000039)	0.000042 (0.000077)	0.00032*** (0.000069)	0.00055*** (0.00011)	0.000076 (0.00021)	0.00012** (0.000057)
DD30	-0.0062*** (0.00022)	-0.0084*** (0.00091)	-0.0059*** (0.00018)	-0.0044*** (0.00041)	-0.0045*** (0.00035)	-0.0037*** (0.00035)	-0.0012 (0.0011)	-0.0034*** (0.00026)
RAIN	-0.0025*** (0.00043)	0.017*** (0.0017)	0.0037*** (0.00055)	-0.0039*** (0.00087)	0.0023*** (0.00085)	0.0059*** (0.0020)	0.0099 (0.0073)	0.00046 (0.00064)
DD10*RAIN	-0.000064 (0.000044)	-0.00030 (0.00026)	-0.00022*** (0.000053)	0.00050*** (0.000087)	-0.00024*** (0.000079)	-0.00030** (0.00015)	0.00031 (0.00070)	-0.00043*** (0.000063)
DD30*RAIN	0.0094*** (0.0011)	0.0077 (0.0055)	0.0090*** (0.00088)	0.0027* (0.0016)	-0.0011 (0.0011)	-0.0015 (0.0015)	-0.0089 (0.0078)	0.0052*** (0.0011)
Observations	4,181	1,179	8,295	1,293	3,588	1,133	97	1,988
R-squared	0.529	0.435	0.607	0.652	0.556	0.563	0.750	0.552

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Dependent variable: log yield. Specifications include county fixed effects and linear time trend

TABLE 9. Barley yield regressions

VARIABLES	(1) region 1	(2) region 2	(3) region 3	(4) region 4	(5) region 5	(6) region 6	(7) region 7	(8) region 8
DD10	-0.00021** (0.000097)	-0.00029*** (0.000036)	-0.000050 (0.000055)	-0.00089*** (0.00010)	-0.00022* (0.00011)	-0.00076*** (0.000066)	-0.00046*** (0.000070)	-0.00023*** (0.000065)
DD30	-0.0066*** (0.0011)	-0.0017*** (0.00056)	-0.016*** (0.00066)	0.0045*** (0.00053)	-0.0033** (0.0013)	0.0022*** (0.00053)	0.00076** (0.00037)	-0.0034*** (0.00065)
RAIN	0.0020 (0.0016)	-0.0025*** (0.00048)	0.011*** (0.0011)	-0.0026* (0.0015)	0.00076 (0.0015)	-0.0046*** (0.00064)	0.0031** (0.0013)	0.012*** (0.0012)
DD10*RAIN	-0.0014*** (0.00022)	-0.000050 (0.000062)	-0.0016*** (0.00014)	0.00024* (0.00015)	0.000023 (0.00018)	0.000062 (0.000077)	0.00068*** (0.00025)	-0.00014 (0.00025)
DD30*RAIN	-0.033*** (0.0062)	-0.0084*** (0.0031)	0.0097** (0.0045)	-0.0046* (0.0027)	0.037*** (0.0082)	-0.013*** (0.0022)	-0.0068* (0.0037)	-0.045*** (0.012)
Observations	928	3,822	3,617	1,658	442	1,704	2,395	2,917
R-squared	0.630	0.530	0.638	0.545	0.396	0.588	0.784	0.805

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Dependent variable: log yield. Specifications include county fixed effects and linear time trend

TABLE 10. Oats yield regressions

VARIABLES	(1) region 1	(2) region 2	(3) region 3	(4) region 4	(5) region 5	(6) region 6	(7) region 7	(8) region 8	(9) region 9
DD10	-0.00030*** (0.000030)	-0.00032*** (0.000028)	-0.00044*** (0.000056)	-0.00053*** (0.000059)	-0.00053*** (0.000055)	-0.00049*** (0.000061)	-0.00015** (0.000071)	-0.00017** (0.000074)	-0.00082*** (0.00014)
DD30	-0.0068*** (0.00030)	-0.012*** (0.00053)	-0.0075*** (0.00055)	0.00076*** (0.00027)	-0.0020*** (0.00051)	-0.00098*** (0.00037)	-0.00098*** (0.00037)	0.00011 (0.00078)	-0.00049 (0.00077)
RAIN	-0.0029*** (0.00041)	-0.0064*** (0.00042)	0.0082*** (0.00094)	0.0025*** (0.00092)	-0.0015* (0.00082)	0.0032*** (0.00072)	-0.0014 (0.00092)	0.0061*** (0.0014)	-0.014*** (0.0018)
DD10*RAIN	-0.00029*** (0.000047)	0.00041*** (0.000051)	-0.00039*** (0.00012)	-0.00027*** (0.000087)	-2.12e-06 (0.000089)	-0.00038*** (0.000068)	0.00010 (0.00011)	0.00072** (0.00028)	0.00063*** (0.00018)
DD30*RAIN	-0.0045** (0.0018)	-0.017*** (0.0025)	-0.013*** (0.0037)	-0.0056*** (0.0012)	-0.0031 (0.0027)	-0.0013 (0.0011)	-0.0055*** (0.0019)	-0.00036 (0.015)	0.0041 (0.0043)
Observations	8,086	6,950	4,133	5,111	1,438	4,028	2,647	2,566	234
R-squared	0.491	0.529	0.502	0.388	0.505	0.385	0.671	0.684	0.664

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Dependent variable: log yield. Specifications include county fixed effects and linear time trend

TABLE 11. Rice yield regressions

VARIABLES	(1) region 1	(2) region 4	(3) region 5	(4) region 6	(5) region 7	(6) region 8	(7) region 9
DD10	0.00023** (0.00010)	0.00054*** (0.00018)	0.000019 (0.00011)	0.00018 (0.00019)	0.00057*** (0.000052)	0.00060 (0.00039)	0.00021*** (0.000028)
DD30	-0.0017*** (0.00059)	-0.0015** (0.00059)	-0.0010** (0.00047)	-0.0014* (0.00070)	-0.0026*** (0.00024)	-0.0033** (0.0016)	-0.0017*** (0.00013)
RAIN	-0.00060 (0.0014)	0.0016 (0.0026)	-0.0075*** (0.0012)	-0.00018 (0.0024)	0.0016** (0.00074)	0.0018 (0.0075)	-0.0017*** (0.00032)
DD10*RAIN	-0.000065 (0.00015)	-0.000095 (0.00021)	0.00051*** (0.00012)	-0.000040 (0.00020)	-0.00025*** (0.000057)	0.00072 (0.0017)	0.000040 (0.000030)
DD30*RAIN	0.0044* (0.0023)	0.0039 (0.0027)	-0.0023 (0.0018)	0.0022 (0.0030)	0.00056 (0.0013)	-0.045 (0.12)	0.0014*** (0.00050)
Observations	100	75	226	158	629	25	1,506
R-squared	0.813	0.666	0.772	0.734	0.854	0.250	0.861

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Dependent variable: log yield. Specifications include county fixed effects and linear time trend

TABLE 12. Upland cotton yield regressions

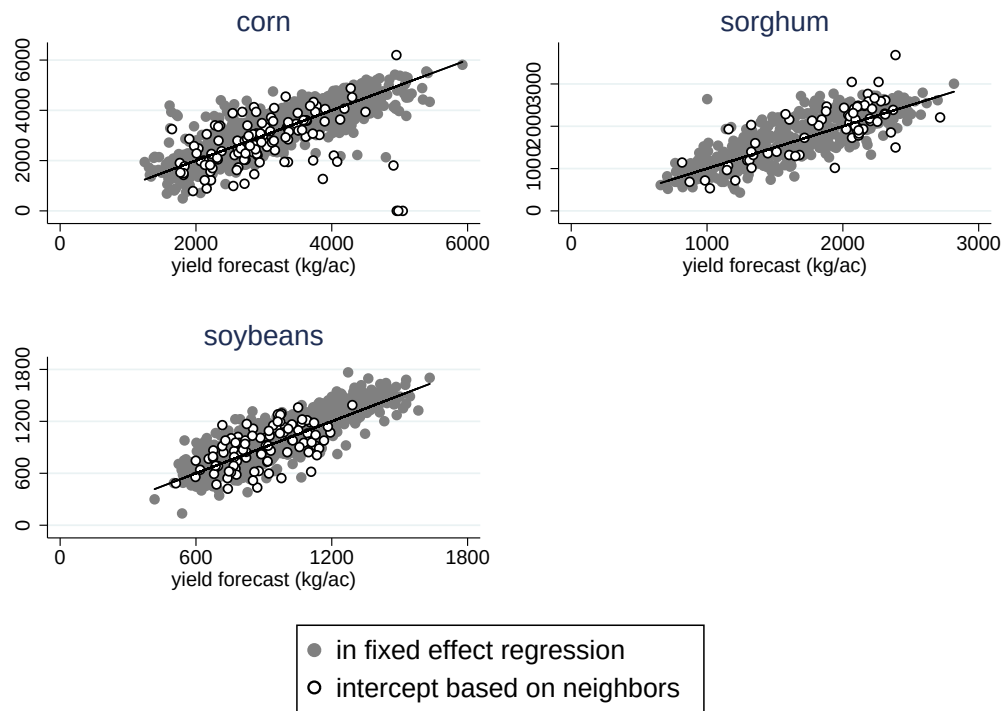
VARIABLES	(1) region 1	(2) region 4	(3) region 5	(4) region 6	(5) region 7	(6) region 8	(7) region 9
DD10	0.00014 (0.00021)	-0.00088*** (0.00010)	0.00047*** (0.00014)	-0.00017*** (0.000058)	0.00068*** (0.000082)	0.00076** (0.00029)	0.00040*** (0.000058)
DD30	-0.0091*** (0.0011)	-0.00050 (0.00041)	-0.013*** (0.0011)	-0.0059*** (0.00034)	-0.0029*** (0.00027)	-0.00095 (0.00096)	-0.0055*** (0.00028)
RAIN	-0.0032 (0.0028)	-0.0059*** (0.0016)	-0.0050*** (0.0016)	0.00036 (0.00077)	-0.0031 (0.0019)	0.0024 (0.016)	-0.00059 (0.00069)
DD10*RAIN	-0.00055** (0.00027)	0.00042*** (0.00016)	0.00019 (0.00017)	0.000099 (0.000070)	0.00026* (0.00015)	-0.00038 (0.0014)	-0.000053 (0.000068)
DD30*RAIN	0.021*** (0.0037)	0.0084*** (0.0019)	0.0047 (0.0054)	0.0053*** (0.0011)	-0.0053*** (0.0013)	0.0031 (0.0076)	0.0050*** (0.0012)
Observations	159	2,991	631	4,420	1,516	106	2,364
R-squared	0.673	0.593	0.541	0.484	0.801	0.616	0.588

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

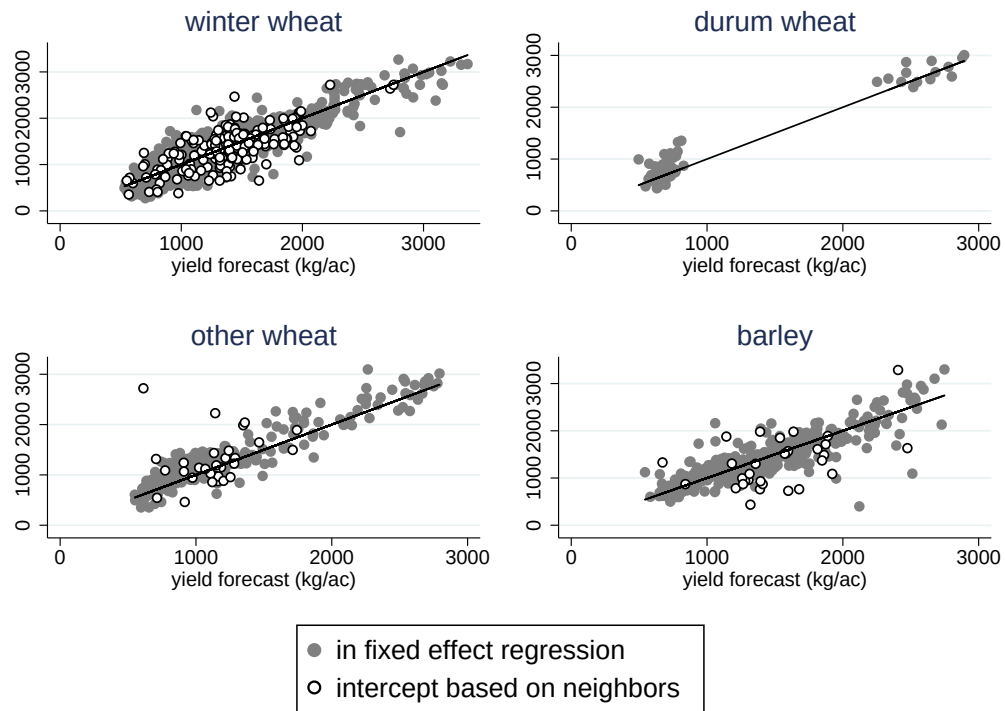
Dependent variable: log yield. Specifications include county fixed effects and linear time trend

FIGURE 1.4. Validation of expected yields, selected crops



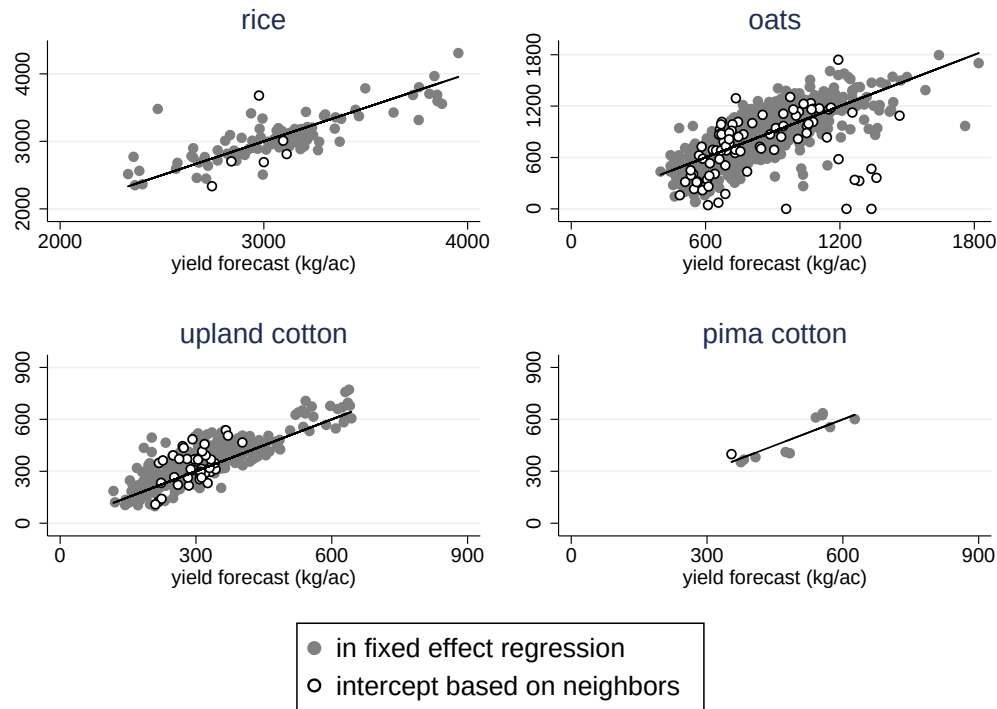
Each point corresponds to a county, with the x-axis indicating expected yield forecasts for that county, averaged over 2006-2011; the y-axis represents the average actual yield for that county, averaged over the same period (reported by the National Agricultural Statistics Service).

FIGURE 1.5. Validation of expected yields, small grains



Each point corresponds to a county, with the x-axis indicating expected yield forecasts for that county, averaged over 2006-2011; the y-axis represents the average actual yield for that county, averaged over the same period (reported by the National Agricultural Statistics Service).

FIGURE 1.6. Validation of expected yields, selected crops



Each point corresponds to a county, with the x-axis indicating expected yield forecasts for that county, averaged over 2006–2011; the y-axis represents the average actual yield for that county, averaged over the same period (reported by the National Agricultural Statistics Service).

References

- Aguirregabiria, Victor and Pedro Mira, “Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models,” *Econometrica*, 2002, 70 (4), 1519–1543.
- Altug, Sumru and Robert A. Miller, “The Effect of Work Experience on Female Wages and Labour Supply,” *The Review of Economic Studies*, 1998, 65 (1), 45–85.
- Arcidiacono, Peter and John Bailey Jones, “Finite Mixture Distributions, Sequential Likelihood and the EM Algorithm,” *Econometrica*, 2003, 71 (3), pp. 933–946.
- _____ and Paul B. Ellickson, “Practical Methods for Estimation of Dynamic Discrete Choice Models,” *Annual Review of Economics*, 2011, 3 (1), 363–394.
- _____ and Robert A. Miller, “Conditional Choice Probability Estimation of Dynamic Discrete Choice Models With Unobserved Heterogeneity,” *Econometrica*, 2011, 79 (6), 1823–1867.
- Arellano, Manuel and Stephen Bond, “Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations,” *The Review of Economic Studies*, 1991, 58 (2), 277–297.

- Barros, Sergio**, “Brazil Annual Biofuels Report,” Technical Report, USDA Foreign Agricultural Service 2012.
- Berry, Steven and Wolfram Schlenker**, “Technical Report for the ICCT: Empirical Evidence on Crop Yield Elasticities,” Technical Report 2011.
- Bhattacharya, Rabi N. and Mukul Majumdar**, “Controlled semi-markov models - the discounted case,” *Journal of Statistical Planning and Inference*, 1989, *21* (3), 365 – 381.
- Boryan, Claire, Zhengwei Yang, Rick Mueller, and Mike Craig**, “Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program,” *Geocarto International*, 2011, *26* (5), 341–358.
- Chomitz, Kenneth M. and David A. Gray**, “Roads, Land Use, and Deforestation: A Spatial Model Applied to Belize,” *The World Bank Economic Review*, 1996, *10* (3), 487–512.
- **and Timothy S. Thomas**, “Determinants of Land Use in Amazonia: A Fine-Scale Spatial Analysis,” *American Journal of Agricultural Economics*, 2003, *85* (4), 1016–1028.
- Claassen, Roger and Abebayehu Tegene**, “AGRICULTURAL LAND USE CHOICE: A DISCRETE CHOICE APPROACH,” *Agricultural and Resource Economics Review*, 1999, *28* (1).
- Conley, Timothy G.**, “{GMM} estimation with cross sectional dependence,” *Journal of Econometrics*, 1999, *92* (1), 1 – 45.
- , *The New Palgrave Dictionary of Economics. Second Edition.*, Palgrave Macmillan, 2008.
- Conlon, Christopher T.**, “A Dynamic Model of Costs and Margins in the LCD TV Industry,” *Working paper*, 2010.
- Deininger, Klaus and Bart Minten**, “Determinants of Deforestation and the Economics of Protection: An Application to Mexico,” *American Journal of Agricultural Economics*, 2002, *84* (4), 943–960.
- Dempster, A. P., N. M. Laird, and D. B. Rubin**, “Maximum Likelihood from Incomplete Data via the EM Algorithm,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 1977, *39* (1), pp. 1–38.
- Fezzi, Carlo and Ian J. Bateman**, “Structural Agricultural Land Use Modeling for Spatial Agro-Environmental Policy Analysis,” *American Journal of Agricultural Economics*, 2011, *93* (4), 1168–1188.
- Flach, Bob, Karin Bendz, and Sabine Lieberz**, “EU Annual Biofuels Report,” Technical Report, USDA Foreign Agricultural Service 2012.
- Gowrisankaran, Gautam and Marc Rysman**, “Dynamics of Consumer Demand for New Durable Goods,” NBER Working Papers 14737, National Bureau of Economic Research, Inc February 2011.
- Hall, Robert E**, “Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence,” *Journal of Political Economy*, December 1978, *86* (6), 971–87.

- Hendel, Igal and Aviv Nevo**, “Measuring the Implications of Sales and Consumer Inventory Behavior,” *Econometrica*, November 2006, *74* (6), 1637–1673.
- Hotz, V Joseph and Robert A Miller**, “Conditional Choice Probabilities and the Estimation of Dynamic Models,” *Review of Economic Studies*, July 1993, *60* (3), 497–529.
- Irwin, Elena G. and Nancy E. Bockstael**, “Interacting agents, spatial externalities and the evolution of residential land use patterns,” *Journal of Economic Geography*, 2002, *2* (1), 31–54.
- Kasahara, Hiroyuki and Katsumi Shimotsu**, “Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices,” *Econometrica*, 2009, *77* (1), pp. 135–175.
- Lin, William, Paul C. Westcott, Robert Skinner, Scott Sanford, and Daniel G. De La Torre Ugarte**, “Supply Response Under the 1996 Farm Act and Implications for the U.S. Field Crops Sector,” Technical Report, USDA Economic Research Service 2000.
- Lubowski, Ruben N.**, “Determinants of Land-Use Transitions in the United States: Econometric Analysis of Changes among the Major Land-Use Categories.” PhD dissertation, Harvard University 2002.
- , **Andrew J. Plantinga, and Robert N. Stavins**, “Land-use change and carbon sinks: Econometric estimation of the carbon sequestration supply function,” *Journal of Environmental Economics and Management*, March 2006, *51* (2), 135–152.
- , ———, and ———, “What Drives Land-Use Change in the United States? A National Analysis of Landowner Decisions,” *Land Economics*, 2008, *84* (4), 529–550.
- Magnac, Thierry and David Thesmar**, “Identifying Dynamic Discrete Decision Processes,” *Econometrica*, 2002, *70* (2), 801–816.
- McLachlan, Geoffrey J. and David Peel**, *Finite Mixture Models*, John Wiley & Sons, 2000.
- Munroe, Darla K., Jane Southworth, and Catherine M. Tucker**, “Modeling Spatially and Temporally Complex Land-Cover Change: The Case of Western Honduras*,” *The Professional Geographer*, 2004, *56* (4), 544–559.
- Murphy, Alvin**, “A Dynamic Model of Housing Supply,” 2012. Working paper.
- Newey, Whitney K. and Kenneth D. West**, “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix,” *Econometrica*, 1987, *55* (3), pp. 703–708.
- Norets, Andriy and Satoru Takahashi**, “On the Surjectivity of the Mapping Between Utilities and Choice Probabilities,” 2012. Working Paper.
- Pakes, Ariel, Michael Ostrovsky, and Steven Berry**, “Simple estimators for the parameters of discrete dynamic games (with entry/exit examples),” *The RAND Journal of Economics*, 2007, *38* (2), 373–399.
- Pesendorfer, Martin and Philipp Schmidt-Dengler**, “Asymptotic Least Squares Estimators for Dynamic Games,” *The Review of Economic Studies*, 2008, *75* (3), 901–928.

- Pinto, Alessandro De and Gerald C. Nelson**, “Land Use Change with Spatially Explicit Data: A Dynamic Approach,” *Environmental and Resource Economics*, 2009, *43*, 209–229. 10.1007/s10640-008-9232-x.
- Roberts, Michael J. and Wolfram Schlenker**, “Identifying Supply and Demand Elasticities of Agricultural Commodities: Implications for the US Ethanol Mandate,” *Forthcoming, American Economic Review*, 2013.
- Rust, John**, “Optimal Replacement of GMC Bus Engines: an Empirical Model of Harold Zurcher,” *Econometrica*, 1987, *55* (5), 999–1033.
- Schlenker, Wolfram and Michael J. Roberts**, “Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change,” *Proceedings of the National Academy of Sciences*, 2009, *106* (37), 15594–15598.
- Searchinger, Timothy, Ralph Heimlich, R. A. Houghton, Fengxia Dong, Amani Elobeid, Jacinto Fabiosa, Simla Tokgoz, Dermot Hayes, and Tun-Hsiang Yu**, “Use of U.S. Croplands for Biofuels Increases Greenhouse Gases Through Emissions from Land-Use Change,” *Science*, 2008, *29*, 1238–1240.
- Souza-Rodrigues, Eduardo A.**, “Demand for Deforestation in the Amazon,” 2012. Working Paper.
- Tyner, Wallace E., Farzad Taheripour, Qianlai Zhuang, Dileep Birur, and Uris Baldos**, “Land Use Changes and Consequent CO2 Emissions due to US Corn Ethanol Production: A Comprehensive Analysis,” Technical Report 2010.
- US EPA**, “Biofuels and the Environment: First Triennial Report to Congress,” Technical Report, United States Environmental Protection Agency 2011.
- , “Regulation of Fuels and Fuel Additives: 2012 Renewable Fuel Standards; Final Rule,” Technical Report, United States Environmental Protection Agency 2012.
- Vance, Colin and Jacqueline Geoghegan**, “Temporal and spatial modelling of tropical deforestation: a survival analysis linking satellite and household survey data,” *Agricultural Economics*, 2002, *27* (3), 317–332.
- Wu, JunJie, Richard M. Adams, Catherine L. Kling, and Katsuya Tanaka**, “From Microlevel Decisions to Landscape Changes: An Assessment of Agricultural Conservation Policies,” *American Journal of Agricultural Economics*, 2004, *86* (1), 26–41.

CHAPTER 2

Indirect Estimation of Yield-Price Elasticities

2.1. Introduction

Although they play a crucial role in determining the environmental effects of changes in agricultural markets, the magnitudes of intensive crop supply responses have weak and largely outdated empirical foundations. While existing studies on yield-price elasticities have relied on regressing realized yields directly on prices, such elasticities may instead be indirectly estimated by estimating input use elasticities and then using economic theory to map to yield elasticities. Indirect estimation delivers considerably more precisely estimated elasticities, and suggests that yield-price elasticities used in influential policy reports are far too high.

Agricultural intensification (yield gains) is the magic bullet when it comes to the trade-off between food production and environmental destruction. Extensive agricultural supply responses – i.e., expansion of agricultural land into natural terrain – has tremendous costs in terms of ecological destruction and greenhouse gas emissions. While intensification is not without environmental costs – e.g. synthetic nitrogenous fertilizers have a substantial carbon footprint – the externalities associated with intensification are generally much smaller than with extensification (Burney et al., 2010).

The relative magnitudes of yield elasticities and acreage elasticities determine whether equilibrium supply responses come primarily from intensification or acreage expansion. Thus, if yield elasticities were larger, more of the supply response would come from intensification, and the environmental impacts of increased food production would be much smaller. Because they make such a big difference in the assessment of environmental impacts, yield elasticities have become a point of contention in the evaluation of biofuels policy. Berry (2011) questioned Tyner et al.'s (2010) use of a yield-price elasticity of .25, arguing that there was no empirical evidence to support a yield elasticity so high. While Houck and Gallagher (1976) found evidence to support yield-price elasticities for US corn as high as .25 and significantly different from zero, subsequent work by Menz and Pardey (1983) and Choi and Helmberger (1993) estimated yield-price elasticities for US corn which were insignificantly different from zero, and with standard errors suggesting yield-price elasticities were unlikely to be larger than .3. More recently, Berry and Schlenker (2011) argue that yield-price elasticities for major US crops are unlikely to be larger than .1.

Theory provides a mapping between fertilizer use elasticities and expected yield elasticities. Consequently, there are two potential approaches to estimating how expected yields respond to price changes: direct estimation of how yields respond to prices, or indirect estimation by estimating the fertilizer use elasticities and using the theoretical mapping.¹ The only economic assumption required for the mapping

¹There have been many studies on fertilizer demand (Griliches, 1959; Burrell, 1989; Denbaly and Vrooomen, 1993; Kaufmann and Snell, 1997; Williamson, 2011), but none of them take the step of relating input use elasticities to yield elasticities. Choi and Helmberger (1993) follow an estimation strategy which is closer to the indirect approach I propose, but their strategy relies

is that farmers choose input levels optimally. When the set of cultivated fields may change, the mapping does not hold strictly in the aggregate, but I show that indirect estimates of yield elasticities with respect to output prices are positively biased (and therefore still useful for estimating upper bounds) under plausible conditions.

From Houck and Gallagher (1976) to Berry and Schlenker (2011), studies on yield elasticities have relied on a direct regression of realized yields onto prices. However, the indirect approach is more precise. Optimization implies that *expected* yields and fertilizer use are directly proportional, and therefore direct estimation of yield elasticities by regressing expected yields on prices would theoretically have the same level of precision as indirect estimation. However, expected yields are not realistically observable; in practice, direct estimation involves regressing *realized* yields on prices. As in all regressions, the precision of the estimates depends in part on the variability of the error term, and unpredictable variation in weather leads to a relatively large variance in the error term in a direct yield-price regression. Indirect estimation avoids this source of noise.

Precision is not the only reason for revisiting the topic of yield-price elasticities. With the exception of Berry and Schlenker (2011), all of the studies of yield elasticities cited above are based entirely on pre-1990 data, typically including data beginning in the 1950's or 60's and extending through the 70's. One concern with estimates based on such old data is that technological change may have changed yield elasticities. Another serious concern comes from the fact that United States agriculture was in a state of transition during the 1960's and early 70's, and it was not until later in the 70's that the use of synthetic fertilizer was pervasive.² Indeed, Menz and Pardey (1983) find evidence of a structural shift in fertilizer usage patterns in the early 1970's. Thus, older estimates may conflate the decision of how much fertilizer to use with the decision of whether or not to adopt synthetic fertilizer, whereas only the "how much" decision is of practical relevance in the United States today.

It should be noted that much of the agronomic literature on fertilizer use and yields focuses on the question of what the optimal level of fertilizer use is (e.g., Cerrato and Blackmer (1990); Johnson and Raun (2003)). The uncertainties in the optimal level of fertilizer use calls into question a basic assumption of the indirect estimation approach: that farmers know the production functions for their fields and maximize profits accordingly. Despite this, the assumption of optimal input decisions has been popular in the agricultural economics literature (e.g., Houck and Gallagher (1976); Choi and Helmberger (1993)). While a more flexible

on additionally estimating how yields depend on input levels, an estimation which suffers from the same problem as the direct approach – randomness in yields introduced by weather makes it relatively less precise.

²According to the 2007 US Census of Agriculture, 266 million acres were treated with commercial fertilizers while 310 acres of cropland were harvested. In 1964, 151 million acres were fertilized, and 287 million acres were harvested. According to the Economic Research Service, Nitrogen applied to corn increased from 58 pounds per acre in 1964 to 135-140 pounds per acre in recent years. By the late 1970's, farmers were already applying over 130 pounds per acre.

model of farmer behavior which does not rely on optimization and perfect information could be estimated in principle, estimating a flexible model along such lines would probably require detailed micro data on farmer behavior, and the lack of such data has been a constraint in the literature on intensive crop supply responses to date.

My indirect estimates suggest that the yield elasticity with respect to output price is unlikely to be larger than .04 for US corn, .11 for soybeans, and .13 wheat. In contrast, direct estimation cannot rule out yield-output-price elasticities as high as .27 for corn, .46 for soybeans, and .19 for wheat. Point estimates of yield-price elasticities are insignificantly different from zero in all cases, and so they may be considerably closer to zero than the upper bounds suggest.

This paper focuses how crop yields respond to price changes through the channel of fertilizer use decisions – i.e., short-run yield-price responses. Longer term yield-price responses – e.g., endogenous farm capital investment or technological change – are beyond the scope of this paper.

Section 2.2 develops the theory behind the indirect estimation approach. Section 2.3 describes the data and specific regressions that I estimate. Section 3.3 presents the results, and Section 2.5 concludes.

2.2. Theory

2.2.1. One input. A farmer growing corn chooses the fertilizer application rate L to maximize expected profits:

$$(2.1) \quad L^*(P_Y, P_L) = \arg \max_L \{P_C Y(L) - P_L L\}$$

where P_Y is the expected price of the output, P_L is the price of fertilizer, and $Y(L)$ is the expected yield at input level L . It should be emphasized that the farmer's decision is made with respect to *expected* yields and prices – fertilizer input decisions are typically made shortly before or after planting, before unpredictable weather events during the growing season determine the realized yields.

The first-order condition for the optimal input choice L^* is

$$(2.2) \quad \frac{dY(L^*)}{dL} = \frac{P_L}{P_Y}.$$

Thus, the optimal fertilizer application rate depends only on the ratio of input and output prices, and we can write $L^*(P_r)$ where $P_r = P_L/P_Y$.

Applying the chain rule, the price ratio ends up relating the yield derivative and the input use derivative:³

$$(2.3) \quad \frac{dY(L^*)}{dP_r} = \frac{dY(L^*)}{dL} \frac{dL^*}{dP_r} = P_r \frac{dL^*}{dP_r}.$$

Applying the implicit function theorem to (2.2), we can see that the yield derivative and fertilizer use derivative with respect to price are both proportional to the second derivative of the production function:

$$\begin{aligned} \frac{dY(L^*)}{dP_r} &= P_r \left(\frac{d^2Y(L^*)}{dL^2} \right)^{-1} \\ \frac{dL^*}{dP_r} &= \left(\frac{d^2Y(L^*)}{dL^2} \right)^{-1}. \end{aligned}$$

It is this proportionality between expected yield and input use elasticities that allows for indirect estimation.

Equation (2.3) can also be expressed in the form of elasticities:

$$(2.4) \quad \mathcal{E}_{Y,P_r} = P_r \frac{L^*}{Y^*} \mathcal{E}_{L,P_r} = \frac{P_L L^*}{P_Y Y^*} \mathcal{E}_{L,P_r}$$

where $\mathcal{E}_{L,P_r} \equiv \frac{P_r}{L^*} \frac{dL^*}{dP_r}$ and $\mathcal{E}_{Y,P_r} \equiv \frac{P_r}{Y^*} \frac{dY^*}{dP_r}$.

Thus, assuming that farmers' decisions satisfy the first-order condition (2.2), one may calculate the yield elasticity using the fertilizer-use elasticity, input expenditures, and expected revenues.

2.2.2. Aggregation. Given that all farmers face the same prices and all choose inputs optimally, the relationship between yield and fertilizer use elasticities holds in the aggregate for a fixed group of active farms, even if those farms have different production functions. Given that equations (2.3) and (2.3) hold for each farm, the equations can be summed or averaged across farms and will still hold. Thus, the mapping from input use to yield elasticities holds in the aggregate, with the input level L^* referring to the average input use across farms.

However, the potential pitfall with aggregation comes from the fact that the set of fields in a given crop may change over time. For example, if the set of fields in corn is growing, then aggregate changes in nitrogen fertilizer used for corn production will reflect not only changes in the optimal usage levels for incumbent corn fields, but will also include fertilizer used on new corn fields. In other words, equations (2.3) and (2.4) do not apply to newly planted fields because input use is not at an interior solution, and therefore we cannot take the derivative of yields with respect to input use for such fields.

I argue that, fortunately, acreage responses are likely to introduce positive bias into indirect estimates of aggregate of yield-price elasticities, meaning that indirect estimation still provides a useful tool for estimating upper bounds on yield-price elasticities.

³Technically, we must assume that $Y(L)$ is concave, twice continuously differentiable and that $L^*(P_r)$ is at an interior solution to guarantee that $\frac{dL^*}{dP_r}$ is well defined.

To make this argument formally, I must expand the notation. Suppose average yields are defined as follows:

$$\bar{Y}(P_r) \equiv \frac{\int_0^{A^*(P_r)} Y_i^*(P_r) di}{A^*(P_r)}$$

where i indexes fields, A_i is the acreage of field i , $Y_i^*(P_r)$ are the optimal yields for field i with price ratio P_r , and $A^*(P_r)$ is the equilibrium acreage with all fields such that $i \leq A^*(P_r)$ being planted.

The aggregate yield elasticity is the elasticity of $\bar{Y}(P_r)$ with respect to P_r . Combined, an average yield elasticity and an acreage elasticity imply a supply elasticity. For many purposes, knowing the yield elasticities for individual fields would be ideal, but for applications depending on supply elasticities, knowing average yield elasticities is sufficient.

Define average fertilizer use similarly:

$$\bar{L}(P_r) \equiv \frac{\int_0^{A^*(P_r)} L_i^*(P_r) di}{A^*(P_r)}$$

where $L_i^*(P_r)$ is the optimal level of per-acre fertilizer use in field i (conditional on the field's being planted).

We can now restate formally the difficulty with aggregation. While optimal input use implies

$$\forall i : \quad \frac{dY_i^*(P_r)}{dP_r} = P_r \frac{L_i^*(P_r)}{dP_r},$$

the aggregate relationship does not hold:

$$\frac{d\bar{Y}(P_r)}{dP_r} \neq P_r \frac{d\bar{L}(P_r)}{dP_r}$$

because of the endogeneity of $A^*(P_r)$. However, I argue that the indirect aggregate estimate provides a lower bound for the aggregate yield elasticity with respect to P_r (or an upper bound on the yield elasticity with respect to the output price) given a plausible assumption.

PROPOSITION 1. *Assuming that marginal fields are no more profitable than the average cultivated field in terms of expected revenues net of fertilizer costs,*

$$\mathcal{E}_{\bar{Y}, P_r} \geq P_r \frac{\bar{L}(P_r)}{\bar{Y}(P_r)} \mathcal{E}_{\bar{L}, P_r}.$$

Furthermore, since $\mathcal{E}_{\bar{Y}, P_Y} = -\mathcal{E}_{\bar{Y}, P_r}$,

$$\mathcal{E}_{\bar{Y}, P_Y} \leq -P_r \frac{\bar{L}(P_r)}{\bar{Y}(P_r)} \mathcal{E}_{\bar{L}, P_r}.$$

While it is theoretically possible for the revenues net of fertilizer costs for marginal fields to be higher than fields which are already planted, it would be extremely surprising.

PROOF. Using Leibniz's rule, the aggregate yield elasticity can be written:

$$(2.5) \quad \mathcal{E}_{\bar{Y}, P_r} = \frac{P_r}{\bar{Y}(P_r)} \frac{\int_0^{A^*(P_r)} \frac{dY_i^*(P_r)}{dP_r} di}{A^*(P_r)} + \mathcal{E}_{A^*, P_r} \left(\frac{Y_{A^*(P_r)}^*(P_r) - \bar{Y}(P_r)}{\bar{Y}(P_r)} \right).$$

where $Y_{A^*(P_r)}^*(P_r)$ denotes the yield of the marginal field. A similar equation holds for the aggregate fertilizer elasticity:

$$(2.6) \quad \mathcal{E}_{\bar{L}, P_r} = \frac{P_r}{\bar{L}(P_r)} \frac{\int_0^{A^*(P_r)} \frac{dL_i(P_r)}{dP_r} di}{A^*(P_r)} + \mathcal{E}_{A^*, P_r} \left(\frac{L_{A^*(P_r)}^*(P_r) - \bar{L}(P_r)}{\bar{L}(P_r)} \right).$$

Multiplying equation (2.6) by $P_r \frac{\bar{L}(P_r)}{\bar{Y}(P_r)}$ and rearranging,

$$(2.7) \quad \frac{P_r}{\bar{Y}(P_r)} \frac{\int_0^{A^*(P_r)} \frac{dY_i^*(P_r)}{dP_r} di}{A^*(P_r)} = P_r \frac{\bar{L}(P_r)}{\bar{Y}(P_r)} \mathcal{E}_{\bar{L}, P_r} - P_r \frac{\bar{L}(P_r)}{\bar{Y}(P_r)} \mathcal{E}_{A^*, P_r} \left(\frac{L_{A^*(P_r)}^*(P_r) - \bar{L}(P_r)}{\bar{L}(P_r)} \right).$$

Finally, substituting equation (2.7) into equation (2.5),

$$\mathcal{E}_{\bar{Y}, P_r} = P_r \frac{\bar{L}(P_r)}{\bar{Y}(P_r)} \mathcal{E}_{\bar{L}, P_r} + \mathcal{E}_{A^*, P_r} \left(\frac{Y_{A^*(P_r)}^*(P_r) - \bar{Y}(P_r)}{\bar{Y}(P_r)} - P_r \left(\frac{L_{A^*(P_r)}^*(P_r) - \bar{L}(P_r)}{\bar{Y}(P_r)} \right) \right).$$

Given that the acreage-price elasticity is negative, the bias of the indirect estimate $P_r \frac{\bar{L}(P_r)}{\bar{Y}(P_r)} \mathcal{E}_{\bar{L}, P_r}$ has the same sign as

$$P_Y \left(Y_{A^*(P_r)}^*(P_r) - \bar{Y}(P_r) \right) - P_L \left(L_{A^*(P_r)}^*(P_r) - \bar{L}(P_r) \right),$$

which is precisely how the profits of a marginal field differ from the average cultivated field (accounting only for the costs of the modeled input). Given the assumption, the difference in profits is negative, so the bias in the indirect estimate of the yield elasticity with respect to P_r is also negative. Conversely, $-P_r \frac{\bar{L}(P_r)}{\bar{Y}(P_r)} \mathcal{E}_{\bar{L}, P_r}$ has positive bias as an estimate of the yield elasticity with respect to P_Y . $\square$

2.2.3. Multiple inputs. This theory extends to the case of multiple inputs. Let $\mathbf{L} = (L_1, \dots, L_J)$ be a vector of J inputs, and $\mathbf{P} = (P_1/P_Y, \dots, P_J/P_Y)$ be the vector of input-output price ratios. The first-order condition (2.2) becomes

$$(2.8) \quad \nabla_{\mathbf{L}} Y = \mathbf{P}.$$

The derivative condition (2.3) becomes

$$(2.9) \quad \nabla_{\mathbf{P}} Y^* = (J_{\mathbf{P}} \mathbf{L}^*)' \mathbf{P}.$$

where $J_{\mathbf{P}}\mathbf{L}^*$ is the Jacobian of the optimal input choice vector $\mathbf{L}^*$ with respect to $\mathbf{P}$. The elasticity condition (2.4) becomes

$$(2.10) \quad \mathcal{E}_{Y,\mathbf{P}} = \mathcal{E}'_{\mathbf{L},\mathbf{P}} X(\mathbf{P})$$

where $X(\mathbf{P}) = \left(\frac{P_1^* L_1}{P_Y Y^*}, \dots, \frac{P_J^* L_J}{P_Y Y^*} \right)$ is a vector containing the ratio of expenditure to expected revenue for each input, and $\mathcal{E}_{\mathbf{L},\mathbf{P}}$ is the matrix of input use-price elasticities; i.e.,

$$\mathcal{E}_{\mathbf{L},\mathbf{P}} = \begin{bmatrix} \mathcal{E}_{L_1, P_1/P_Y} & \cdots & \mathcal{E}_{L_1, P_J/P_Y} \\ \vdots & \ddots & \vdots \\ \mathcal{E}_{L_J, P_1/P_Y} & \cdots & \mathcal{E}_{L_J, P_J/P_Y} \end{bmatrix}.$$

Finally, note that the yield elasticity with respect to the output price (holding input prices fixed), can simply be obtained by summing the yield-input price elasticities for each input; i.e.,

$$(2.11) \quad \frac{\partial \ln Y^*}{\partial \ln P_Y} = \sum_{j=1}^J \frac{\partial \ln Y^*}{\partial \ln (P_j/P_Y)} \frac{d \ln (P_j/P_Y)}{d \ln (P_Y)} = - \sum_j \frac{\partial \ln Y^*}{\partial \ln (P_j/P_Y)}$$

since $\frac{d \ln (P_j/P_Y)}{d \ln (P_Y)} = -1$ for all j (given that P_j is fixed).

Finally, the argument behind Proposition 1 extends naturally to the case of multiple inputs, so assuming that the revenues net of total fertilizer costs for the average cultivated field is higher than for marginal fields, we can conclude that

$$\bar{\mathcal{E}}_{Y,P_Y} \leq - \sum_j \sum_{j'} X_{j'}(\mathbf{P}) \bar{\mathcal{E}}_{L_j, P_{j'}/P_Y}.$$

2.3. Data and estimation

In this section, I describe the implementation of the indirect estimation approach for US corn, soybeans, and wheat. Indirect estimation is based on estimates of input use elasticities for nitrogenous fertilizers (abbreviated by N), phosphates (P), and potash (K). For comparison, I also compute direct estimates of yield-price elasticities.

2.3.1. Data. Data on fertilizer use and prices were obtained from the National Agricultural Statistics Service and date back to 1990 (with some missing years). Fertilizer application data is available for each crop in most of the states where the crop is prominent. State-level fertilizer price data is more sparse, so national average prices paid are used when state-level prices are missing. The nitrogen price is derived from

the price of anhydrous ammonia; the phosphate price, from the price of superphosphate 44-46%; the potash price, from the price of muriate of potash 60-62%.⁴

Expected output prices are taken from futures prices obtained from the Chicago Board of Trade. The expected prices for corn and soybeans are the average prices in January for contracts with delivery in the November or December. For winter wheat, the expected price for year t is the average price in the September of year $t - 1$ for contracts with delivery in December of year t .

While realized yields are used in the direct estimation of yield elasticities, indirect estimation calls for a measure of expected yields (see equation (2.4)). Expected yields are computed using the county-level yield forecasts computed by Scott (2013), and then aggregating to the state level by taking an average weighted by harvested acreage.⁵

2.3.2. Empirical models. I begin by estimating yield-price elasticities directly, using the following regression:

$$(2.12) \quad \ln(Y_{st}) = \alpha \ln(P_t) + f(t) + \alpha_{0s} + \varepsilon_{st}$$

where s indexes US states, t indexes years, Y_{st} is the realized yield, P_t is the expected output price, $f(t)$ is a time trend, and α_{0s} is a state fixed effect. I estimate equation (2.12) separately for corn, soybeans, and winter wheat. The parameter α is the direct estimate of the yield-price elasticity. For all regression equations with a time trend, I use a cubic spline with three knots for $f(t)$.

I also estimate the following more flexible direct model:

$$(2.13) \quad \ln(Y_{st}) = \sum_j \alpha_j \ln(P_t/P_j) + f(t) + \alpha_{0s} + \varepsilon_{st}$$

where j indexes inputs, $\sum_j \alpha_j$ is the yield-price elasticity (with respect to the output price, holding input prices fixed), and the inputs included are nitrogen, phosphate, and potash.

Indirect estimation of yield-price elasticities begins with estimates of input use elasticities. I estimate a restricted model of input use elasticities with cross-price elasticities set to zero,

$$(2.14) \quad \ln(L_{jst}) = \gamma_j \ln(P_t/P_j) + f(t) + \alpha_{0js} + \varepsilon_{jst},$$

where L_{jst} is the per-acre input use for nutrient j in state s during year t (by crop).

⁴Prices were converted to prices per nutrient short ton using chemical masses. For example, the price of nitrogen is the price per short ton of ammonia times 17/14. Data on fertilizer use are in terms of nutrient tons, so converting prices in this way is necessary to compute fertilizer expenditure appropriately.

⁵Note that these measures of expected yields could not be plugged into the direct estimation approach, for they are constructed without using input and output prices – they only smooth over technological change and eliminate weather variation.

TABLE 1. Average share of expected revenue spent on nutrient inputs

	Nitrogen	Phosphate	Potash
Corn	0.069	0.045	0.031
Soybeans	0.019	0.059	0.046
Winter wheat	0.083	0.068	0.034

US-wide averages, 1990-2010 based on NASS reports.

I also estimate an input use model with a full set of first-order input use elasticities:

$$(2.15) \quad \ln(L_{jst}) = \sum \gamma_{jj'} \ln(P_t/P_{j'}) + f(t) + \alpha_{0js} + \varepsilon_{jst}.$$

All regression equations are estimated separately for each crop. As a robustness check, I also estimate differenced versions of equations (2.14) and (2.15) which feature differenced input use and price ratios but omit the time trend and fixed effects.

2.3.3. Indirect estimation example. Indirect estimation requires two inputs: estimates of the revenue shares of input expenditures, and estimates of input use elasticities. Table 1 presents revenue shares of input expenditure for each crop and nutrient, based on averages of such shares across states and years (weighted by harvested area). Table 2 presents input elasticities (estimates of equation (2.15)) for corn with a full set of cross-price input use elasticities (equation (2.15)). Together, these two tables present the numbers we need to indirectly compute a yield-price elasticity.

As described in Section 2.2.3, the indirect estimate is computed by multiplying the matrix of input use elasticities by the vector of revenue shares. Using input elasticities from Table 2 and expenditure shares from Table 1,

$$\begin{pmatrix} -.0146 & .1439 & -.0899 \\ -.0840 & .2321 & .0107 \\ -.1247 & .2749 & .1503 \end{pmatrix} \begin{pmatrix} .0689 \\ .0450 \\ .0305 \end{pmatrix} = \begin{pmatrix} -.0086 \\ .0288 \\ -.0011 \end{pmatrix},$$

and then summing the resulting vector delivers a yield-price elasticity of .019.

All indirect yield-price elasticities presented in Table 4 are estimated according to this indirect procedure. Standard errors on the indirect estimates are derived by sampling input use elasticity matrices from their estimated asymptotic distribution, and then constructing a simulated distribution of yield elasticities by computing the indirect estimate for each matrix of sampled input use elasticities.

2.3.4. On endogeneity. Although the literature on fertilizer demand has largely ignored simultaneity problems, it is worth considering whether unobservable factors which shift the fertilizer demand curve could create bias when estimating equation (2.14) or (2.15).

TABLE 2. Input use elasticities for corn

	$\ln(x_{N,corn})$	$\ln(x_{P,corn})$	$\ln(x_{K,corn})$
$\ln(P_{corn}/P_N)$	-0.0146 (0.0329)	-0.0840 (0.0480)	-0.125 (0.0330)
$\ln(P_{corn}/P_P)$	0.144 (0.0936)	0.232 (0.104)	0.27 (0.0791)
$\ln(P_{corn}/P_K)$	-0.0899 (0.0899)	0.0107 (0.0917)	0.150 (0.0686)
Observations	251	251	250

Standard errors with clustering by year in parentheses. Regressions include cubic spline with three knots and state-level dummy variables.

One might argue that shifts in fertilizer prices are likely to be driven almost entirely by exogenous factors shifting the supply of fertilizer. For example, the natural gas price is plausibly the main determinant of the price of ammonia (and other nitrogenous fertilizers, which are all derived from ammonia), and ammonia production accounts for less than 1.5% of natural gas use in the United States. Furthermore, there are many firms in the fertilizer manufacturing industry producing homogeneous chemical products, so demand shifts are not likely to affect markups.⁶

Furthermore, fertilizer application rates in the US have been relatively stable in the US since the late 1970's,⁷ and changes in crop acreage are generally very gradual, so it's not clear that there are any factors which would shift the demand curve for fertilizer substantially in the short run, and fertilizer production is arguably constant returns to scale in the long run, so gradual changes in the demand curve might not affect prices.

On the supply side, there are undoubtedly large sources of variation in costs, at least for the production of nitrogenous fertilizers. The price of natural gas is highly volatile, and natural gas accounts for on the order of 90% of ammonia production costs Yara (2012), and ammonia is the main input for all nitrogenous fertilizers.

Thus, I argue that endogeneity may not be a large concern when estimating fertilizer use elasticities.

TABLE 3. Direct yield elasticity estimates

	$\ln(Y_{corn})$		$\ln(Y_{soybeans})$		$\ln(Y_{wheat})$	
$\ln(P_{crop})$	0.245		0.203		-0.0348	
	(0.140)		(0.122)		(0.105)	
$\ln(P_{crop}/P_N)$		-0.0395		-0.0719		-0.0733
		(0.0762)		(0.0809)		(0.0767)
$\ln(P_{crop}/P_P)$		-0.184		0.0873		0.0803
		(0.147)		(0.177)		(0.149)
$\ln(P_{crop}/P_K)$		0.181		0.110		-0.0523
		(0.0847)		(0.108)		(0.124)
Observations	495	495	410	410	460	460
Yield Elasticity	0.245	-0.0421	0.203	0.125	-0.0348	-0.0453
95% CI	(-0.05,0.54)	(-0.35,0.27)	(-0.05,0.46)	(-0.21,0.46)	(-0.26,0.19)	(-0.36,0.27)

P_{crop} refers to P_{corn} , $P_{soybeans}$, or P_{wheat} , corresponding to the dependent variable. Standard errors with clustering by year in parentheses. Regressions include cubic spline with three knots and state-level dummy variables.

TABLE 4. Indirect yield-price elasticity estimates

Regression specification	Cross-price elasticities	Corn	Soy	Wheat
levels	no	0.021	0.059	0.008
		(0.011,0.031)	(0.017,0.101)	(-0.039,0.056)
levels	yes	0.019	0.036	0.034
		(0.006,0.032)	(-0.001,0.072)	(-0.039,0.107)
differences	no	0.006	-0.011	0.002
		(-0.004,0.016)	(-0.046,0.023)	(-0.124,0.128)
differences	yes	0.009	-0.024	0.031
		(-0.006,0.023)	(-0.058,0.010)	(-0.027,0.089)

95% Confidence intervals with clustering by year in parentheses.

2.4. Results

Table 3 presents direct estimates of yield-price elasticities for all three crops. All 95% confidence intervals for the yield-price elasticity contain zero, and the narrowest confidence interval has an upper bound of .27 for corn, .46 for soybeans, and .19 for wheat.

⁶According to 2011 USGS Minerals Yearbook publications, there were 13 companies actively producing ammonia (activated nitrogen) in the US in 2011, six companies mining phosphate rock, and three companies producing potash. Producers of fertilizer nutrients also face substantial import competition. In recent years, almost 40% of activated nitrogen consumed in the US has been imported, about 10% of phosphate rock, and over 80% of potash. See USGS Minerals Yearbooks for details.

⁷See, for example, USDA "Fertilizer Use and Price" reports, <http://www.ers.usda.gov/data-products/fertilizer-use-and-price.aspx>.

In contrast, indirect estimates presented in Table 4 are considerably more precise. The largest upper bounds for 95% confidence intervals are just over .03 for corn, .1 for soybeans, and .13 for wheat. In all cases, indirect estimation provides unambiguous gains in precision.

Furthermore, my estimates pin down the magnitude of yield-price elasticities considerably more precisely than other estimates in the literature. After Houck and Gallagher (1976), the trend has been towards yield-price elasticities for US corn which are insignificantly different from zero, with standard errors getting smaller. Menz and Pardey (1983) and Choi and Helmberger (1993) were not able to rule out yield-price elasticities as large as .3. Berry and Schlenker's (2011) unpublished results are compatible with yield-price elasticities for US corn as large as .1. My indirect estimates suggests that the yield-price elasticity for US corn is unlikely to be larger than .04.

2.5. Conclusion

Indirect estimation proves to be a useful tool in obtaining relatively precise estimates of yield-price elasticities. Both direct and indirect estimation deliver point estimates of yield-price elasticities which are insignificantly different from zero, but indirect estimation is considerably more precise, providing evidence against high values of yield-price elasticities which would be compatible with the direct estimates. My indirect estimates suggest that yield-price elasticities are quite small for major US crops – probably no greater than .04 for corn, .11 for soybeans, and .13 for wheat.

My results indicate that yield-price elasticities as high as .25 (used in Tyner et al. (2010)) are far too high, at least for the US. Given Scott's (2013) estimates of long run acreage-price elasticities on the order of .4 for US cropland, acreage responses appear to be the dominant component of crop supply response in the long run.

References

- Berry, Steven**, "Biofuels Policy and the Empirical Inputs to GTAP Models," Technical Report 2011.
- and **Wolfram Schlenker**, "Technical Report for the ICCT: Empirical Evidence on Crop Yield Elasticities," Technical Report 2011.
- Burney, Jennifer A., Steven J. Davis, and David B. Lobell**, "Greenhouse gas mitigation by agricultural intensification," *Proceedings of the National Academy of Sciences*, 2010.
- Burrell, A.**, "THE DEMAND FOR FERTILISER IN THE UNITED KINGDOM," *Journal of Agricultural Economics*, 1989, 40 (1), 1–20.
- Cerrato, M. E. and A. M. Blackmer**, "Comparison of Models for Describing Corn Yield Response to Nitrogen Fertilizer," *Agronomy Journal*, 1990, 82.

- Choi, Jung-Sup and Peter G. Helmberger**, “How Sensitive are Crop Yields to Price Changes and Farm Programs?,” *Journal of Agricultural and Applied Economics*, 1993, 25, 237–244.
- Denbaly, Mark and Harry Vroomen**, “Dynamic Fertilizer Nutrient Demands for Corn: A Cointegrated and Error-Correcting System,” *American Journal of Agricultural Economics*, 1993, 75 (1), 203–209.
- Griliches, Zvi**, “Distributed Lags, Disaggregation, and Regional Demand Functions for Fertilizer,” *Journal of Farm Economics*, 1959, 41 (1), pp. 90–102.
- Houck, James P. and Paul W. Gallagher**, “The Price Responsiveness of U.S. Corn Yields,” *American Journal of Agricultural Economics*, 1976, 58 (4), pp. 731–734.
- Johnson, G. V. and W. R. Raun**, “Nitrogen Response Index as a Guide to Fertilizer Management,” *Journal of Plant Nutrition*, 2003, 26 (2), 249–262.
- Kaufmann, Robert K. and Seth E. Snell**, “A Biophysical Model of Corn Yield: Integrating Climatic and Social Determinants,” *American Journal of Agricultural Economics*, 1997, 79 (1), pp. 178–190.
- Menz, Kenneth M. and Philip Pardey**, “Technology and U.S. Corn Yields: Plateaus and Price Responsiveness,” *American Journal of Agricultural Economics*, 1983, 65 (3), pp. 558–562.
- Scott, Paul T.**, “Dynamic Discrete Choice Estimation of Agricultural Land Use,” *Working Paper*, 2013.
- Tyner, Wallace E., Farzad Taheripour, Qianlai Zhuang, Dileep Birur, and Uris Baldos**, “Land Use Changes and Consequent CO2 Emissions due to US Corn Ethanol Production: A Comprehensive Analysis,” Technical Report 2010.
- Williamson, James M.**, “The Role of Information and Prices in the Nitrogen Fertilizer Management Decision: New Evidence from the Agricultural Resource Management Survey,” *Journal of Agricultural and Resource Economics*, 2011, 36 (3), 552–572.
- Yara**, “Yara Fertilizer Industry Handbook,” Technical Report 2012.

CHAPTER 3

Climate-Saving Dietary Change: The Case of Soy Milk

3.1. Introduction

Could dietary change have a dramatic impact on greenhouse gas emissions? Although there is considerable uncertainty in estimates of the greenhouse gas (GHG) emissions embodied in different foods, it seems like the answer to this question is yes. For example, most studies suggest that beef has a much higher carbon footprint than chicken, regardless of whether carbon footprints measured in terms of weight, calories, or protein content.¹ Nutritionally comparable foods with different carbon footprints may exist, but one could argue that this fact in itself has little normative relevance, for consumers may or may not view such foods as substitutes in practice.

A similar point could be made about whole diets. Many have noted that a shift towards plant-based diets could reduce greenhouse gas emissions, but formal analysis of the issue has largely only evaluated the GHG emissions embodied in different diets while ignoring people's willingness to undergo dietary change. Such studies include Eshel and Martin (2006) and Carlsson-Kanyama and Gonzalez (2009).

How would diets change if food prices reflected the social costs of GHG emissions? Answering this question requires an understanding of how food demand responds to prices. While estimating demand patterns for all food products presents a daunting task, I attempt to shed some light on the issue by focusing on milk consumption.

One reason for focusing on the milk market is that it includes a traditional animal-based product, cow's milk, as well as plant-based products, such as soy milk. Furthermore, the dairy industry merits some attention in its own right, for it is responsible for an estimated 4% of anthropogenic greenhouse gas emissions (FAO, 2010).

I simulate price changes on milk products corresponding to the social costs of embodied GHG emissions, finding very little substitution from cow's milk to soy milk. Overall decreases in milk consumption may be more substantial, but my partial equilibrium model of milk consumption cannot predict how decreased milk consumption will affect net GHG emissions.

Turning to a broader data set, I document a steady positive trend in US soy milk consumption between 2001 and 2006. This trend cannot be explained by price changes, so it calls for an explanation which goes beyond a static demand model. I present evidence which loosely corroborates an explanation in which consumers were uncertain about their preferences for soy milk and had to discover their preferences through experimentation. However, this does not rule out other factors such as consumption externalities or changing perceptions of the health impacts of various types of milk.

¹See Pelletier (2008) and Pelletier et al. (2010) for life cycle assessments of chicken and beef in the US, or Carlsson-Kanyama and Gonzalez (2009) for an assessment of a variety of products in Sweden

Section 3.2 explains the model and estimation strategy, with an econometric strategy following in the tradition of Berry et al. (1995). Section 3.3 presents results and counterfactuals. Section 3.4 documents the growth in soy milk consumption and explores the role of soy milk promotions and discounts.

3.2. Model and estimation

Cow's milk is a commodity when it leaves the farm, but it is a highly differentiated product by the time it reaches consumers. Retailers in the US generally carry private label milks alongside national brands. Cow's milk comes in several levels of fat content as well as organic, flavored, and lactose-free varieties. Non-dairy milks such as soy milk (along with soy-based foods more broadly) have grown dramatically in popularity over the last 15 years, and soy milk drinkers now enjoy a similar degree of product differentiation to cow's milk drinkers.

A discrete choice model with random coefficients allows us to flexibly model demand within this differentiated product market.² Let consumer i 's utility of purchasing product j at store m in month t be denoted by

$$(3.1) \quad u_{ijmt} = \beta_i x_{jmt} + \xi_{jmt} + \varepsilon_{ijmt}$$

where x_{jt} are the characteristics of product j in market t , ξ_{jmt} is a market-level demand shock, and ε_{ijmt} is an idiosyncratic shock, assumed to be identically and independently distributed across i , j , and t according to a type 1 extreme value distribution.

I estimate the model using scanner data from a major US grocery retail chain.³ The data include monthly sales of all cow's milk, soy milk, and rice milk products for 223 Northern California stores between October 2004 and October 2005. The vector of characteristics x includes price, an organic dummy, a nondairy dummy, and lactose-free dummy, and brand dummies. A couple notes about this set of characteristics are in order. First, the dairy and no-lactose dummies are not collinear because some cow's milk products have had the lactose removed, e.g., Lactaid. Furthermore, there is no substantive distinction between soy milk and rice milk in the model – they simply appear as different brands of non-dairy milk. Since there is only one brand of rice milk (Rice Dream) available in the data set, I cannot identify a coefficient on a rice-milk dummy apart from the Rice Dream brand dummy. Table 1 presents summary statistics of the brands in the data.

For each store, I construct a measure of market size based on store area by regressing the average volume of sales on store area and multiplying fitted values for each store by three. Like most measures of market size in the empirical discrete choice literature, this is admittedly quite arbitrary, based on little more than

²See Berry (1994) for an explanation of how the random coefficients model compares to other discrete choice models, or (Train's 2009) chapter on "Mixed Logit" models for a textbook treatment.

³See the SIEPR-Giannini website for more information about the data center: <http://are.berkeley.edu/SGDC/>.

Table 1: Product summary

Brand	organic	non-dairy	lactose-free	avg price	availability	revenue share
Private label A	no	no	no	1.25	1.00	0.420
Private label B	no	no	both	2.08	1.00	0.400
Horizon	yes	no	no	4.06	1.00	0.060
Silk	both	yes	yes	3.78	1.00	0.034
Lactaid	no	no	yes	3.86	1.00	0.030
Private label C	yes	both	both	3.23	1.00	0.030
Nestle	no	no	no	3.97	0.99	0.004
8th Continent	no	yes	yes	3.57	0.97	0.010
Rice Dream	no	yes	yes	3.94	0.97	0.003
Stremicks	yes	no	no	3.47	0.66	0.010
Soy Dream	no	yes	yes	4.11	0.05	0.001
Humbolt Creamery	no	no	no	2.42	0.02	0.140

the intuition that it would probably be very hard to get Americans to drink more than three times as much milk as they normally do, even with extremely low prices.

Another benefit of using a random coefficients model is that it allows us to incorporate demographic data. I allow for correlation between the random coefficients, age, and ethno-racial categories; the joint distribution of these demographic variables by zip code comes from the US Census of Population in 2000. To capture the effects of differential rates of lactose intolerance, I estimate different coefficients on the lactose-free dummy estimated for each of the following ethno-racial categories: Asian, black, white non-Hispanic, and other. I also allow young-adults to have a different mean value for the non-dairy coefficient.

Formally, let v_i represent the demographic characteristics of individual i . For non-price, characteristics, I assume that random coefficients are distributed normally, conditional on demographic characteristics,

$$\beta_i \sim \mathcal{N}(\bar{\beta} + \Gamma(v_i), \Sigma).$$

and the following specification for Γ ,

$$\begin{pmatrix} \Gamma_{non-dairy}(v_i) \\ \Gamma_{lactose-free}(v_i) \end{pmatrix} = \begin{pmatrix} \gamma_a \cdot D_{i,18 \leq AGE_i \leq 30} \\ \sum_r \gamma_r D_{i,r} \end{pmatrix}$$

where $D_{i,18 \leq a \leq 30}$ is an indicator for whether an individual's age is between 18 and 30, and $D_{r,i}$ represents racial dummy variables, where the racial categories are Asian, black, non-Hispanic white, and other.

Furthermore, I restrict Σ such that only the non-dairy coefficient has a purely stochastic component – i.e., the only nonzero element of Σ corresponds to the variance of the non-dairy coefficient for a given value of v_i . While this allows for flexible substitution between dairy and non-dairy milks, the model still places severe restrictions on the cross-price elasticities between products within the non-dairy and dairy

categories. However, within-category substitution patterns are not terribly important for the counterfactuals of interest.

For comparison's sake, I also estimate basic logit and nested logit models. In the nested logit model, products are grouped by the non-dairy dummy, which means the nested logit model is very similar to the random coefficients model, differing only in distributional assumptions.

I estimate θ using a method of moments procedure based on

$$E \left[\hat{\xi}_{jmt}(\theta) z_{jmt} \right] = 0$$

where z_{jmt} is a vector of instruments explained below. I use the the contraction mapping introduced by Berry et al. (1995) to recover demand shocks for a given parameter vector, $\hat{\xi}_{jt}(\theta)$.

First, all non-price product characteristics are included in z – these instruments are very standard. I also construct two instruments representing the number of available products with the same value of the non-dairy variable. For all non-dairy milks, the non-dairy instrument represents the number of other non-dairy milk products available in the market; for dairy products, this instrument is fixed at zero. The second instrument is constructed analogously for dairy milk products. Note that these instruments are a subset of the instruments suggested by Berry et al. (1995) (BLP instruments), and they constitute a minimal set of instruments which, on their own, provide strong identification in the nested logit model.

Next, I include a set of instruments which are interactions of product characteristics with demographic statistics. There is one such instrument for each parameter of Γ in the model. Corresponding to γ_a , I interact the share of young adults in market m with the non-dairy dummy variable. Corresponding to each γ_r , I interact the lactose-free dummy with the share of ethno-racial category r within market m . While it is possible in principle to identify the parameters of Γ with a larger set of BLP instruments, these population-based instruments seem to provide much stronger identification.

Finally, z includes a unit cost measure provided by the retailer. It represents the retailer's estimate of the cost of replacing the item on the shelf. While wholesale prices should not be correlated with local demand shocks, one concern is that the ξ 's could include multi-market components which would be correlated with the unit cost measure.

3.3. Results

Parameter estimates and elasticities. Table 2 displays parameter estimates for the basic logit and random coefficients model. Nested logit coefficient estimates are omitted since the parameter values are not directly comparable.

Table 2: Coefficient estimates				
	RC-Full	RC-Simple	Logit-GMM	Logit-OLS
β_{price}	-0.469 (0.0071)	-0.475 (0.0075)	-0.46 (0.007)	-0.285 (0.0059)
$\sigma_{non-dairy}$	3.68 (0.39)	4.49 (0.55)		
$\gamma_{young,non-dairy}$	1.37 (0.12)			
$\gamma_{asian,no-lactose}$	0.346 (0.051)			
$\gamma_{black,no-lactose}$	1.03 (0.071)			
$\gamma_{other,no-lactose}$	-0.195 (0.1)			
Standard errors in parentheses				

Logit-OLS					Logit-GMM				
Price	Quantity				Price	Quantity			
	PL-A	Horizon	8thC	Silk		PL-A	Horizon	8thC	Silk
PL-A	-0.566	0.443	0.393	0.412	PL-A	-0.903	0.708	0.628	0.658
Horizon	0.006	-2.240	0.017	0.019	Horizon	0.010	-3.550	0.027	0.030
8thC	0.001	0.004	-1.990	0.003	8thC	0.002	0.006	-3.160	0.005
Silk	0.003	0.011	0.010	-2.090	Silk	0.005	0.017	0.015	-3.320

Nested Logit					RC-Full				
Price	Quantity				Price	Quantity			
	PL-A	Horizon	8thC	Silk		PL-A	Horizon	8thC	Silk
PL-A	-0.735	1.640	0.325	0.341	PL-A	-0.919	0.724	0.417	0.438
Horizon	0.022	-3.810	0.014	0.016	Horizon	0.010	-3.620	0.018	0.020
8thC	0.001	0.003	-3.100	0.381	8thC	0.001	0.004	-3.020	0.235
Silk	0.003	0.009	1.080	-2.570	Silk	0.004	0.012	0.666	-2.750

FIGURE 3.1. Mean cross-price elasticities for selected products and specifications.

Unsurprisingly, controlling for price endogeneity leads to more elastic demand, and I find substantial heterogeneity in tastes for non-dairy milk. Furthermore, the ethno-racial coefficient estimates on the lactose-free dummy seem reasonable, given that Asians and blacks are known to have high rates of lactose intolerance (note that non-Hispanic whites are the omitted group).

Figure 1 presents average cross-price elasticity estimates (weighted by market size) for four brands and four model specifications.

As the difference between the Logit-OLS and Logit-GMM specifications makes clear, own-price elasticities are much larger in magnitude after controlling for price endogeneity. However, the own-price elasticities are relatively similar across specifications which do account for price endogeneity.

The differences in cross-price elasticity estimates are more interesting. In the nested logit and random coefficients specifications, the cross-price elasticities between the two soy milk products (8th Continent and Silk) are dramatically higher, highlighting the importance of controlling for heterogeneity.

The estimates suggest that there is very little substitution between cow's milk and soy milk in response to price changes. As a result, the main effect of a price increase for cow's milk products would be to decrease cow's milk consumption, without having much impact on soy milk consumption. Below, I will demonstrate this point formally with counterfactuals.

The own-price elasticity of Private label A merits further comment. Inelastic own-price elasticity estimates sometimes reflect a problem, since locally inelastic demand is inconsistent with optimal price setting in a single-product context. However, the grocery store's price setting problem is a multi-product one. If other non-milk products have negative cross-price elasticities with milk, then the optimal pricing point could very well involve an inelastic own-price elasticity for milk. This story is consistent with industry anecdotes in which milk serves as a loss leader, i.e., to bring shoppers into the grocery store. Note that Private label A is the product with the largest market share and lowest price.

Counterfactuals. *Would milk consumption change dramatically if prices reflected social costs of GHG emissions?* To answer this question, we must first quantify the social costs of GHG emissions embodied in milk.

There have been several studies estimating the carbon footprint of cow's milk. Thoma et al. (2010) estimate that, in the US, the cradle-to-retail production chain for cow's milk involves about 1.8 kilograms of CO₂-equivalent emissions per kilogram of milk, on average. A global analysis of the dairy industry by the Food and Agricultural Organization puts the figure for North America at 1.3 kg of emissions per kilogram of milk (FAO, 2010). It's worth noting that the FAO report also documents substantial regional heterogeneity in milk's carbon footprint, where a gallon of milk produced in some parts of the world embody as much as five times the emissions of a gallon of milk produced in the US.

Soy milk's carbon footprint has not been studied as carefully, at least not in well-documented, published studies. Silk's promotional website claims that production of their soy milk "generates 57% fewer greenhouse gasses than producing a half-gallon of typical dairy milk."⁴ One assessment of tofu's carbon footprint suggests that tofu embodies around 0.86kg CO₂e per kg of tofu.⁵ Since tofu is made from soy milk (and no protein is

⁴From <http://switchtosilk.com/silkvsmilk.php?id=friendly-footprint> on Jan 19, 2012.

⁵From <http://tofuscarbonfootprint.weebly.com/>, visited Jan 19, 2012.

Table 3: Counterfactuals

ΔP		Logit-GMM model		RC-Full model	
dairy	non-dairy	ΔQ , cow	ΔQ , soy	ΔQ , cow	ΔQ , soy
.07	.02	-2.08 (-2.13, -2.02)	0.182 (0.178, 0.187)	-2.11 (-2.17, -2.06)	0.116 (0.0994, 0.129)
.20	.07	-5.86 (-6.03, -5.71)	-0.121 (-0.127, -0.116)	-5.98 (-6.14, -5.82)	-0.0923 (-0.102, -0.0849)
.61	.20	-17.3 (-17.8, -16.8)	-0.285 (-0.316, -0.257)	-17.6 (-18.1, -17.2)	-0.237 (-0.266, -0.214)
1	0	-27.4 (-28.1, -26.7)	14.6 (14.3, 15.)	-27.9 (-28.6, -27.2)	9.22 (7.95, 10.3)
Price changes are measured in dollars, quantity changes in percentages. Estimated 95% confidence intervals in parentheses.					

added in the process), soy milk's carbon footprint per gram of protein should be no larger than tofu's. Given that tofu is approximately 8% protein by weight and soy milk is approximately 3%, this suggests that soy milk's carbon footprint is on the order of 0.3-0.4 kg CO₂e. These numbers suggest that soy milk's per-kg carbon footprint is about one third of cow's milk.

To put a dollar value on the emissions embodied in various products requires a value for the social cost of GHG emissions, and there is little agreement regarding this value. Rather than taking a stand on this parameter, I will simply compute several counterfactuals which correspond to taxing milks in proportion to different carbon prices.

In the following counterfactuals, I simulate price increases on cow's milk and soy milk products. Strictly speaking, actual taxes might not lead to these price changes in equilibrium, for taxes might not fully pass through to consumer prices because of imperfect competition and economies of scale.

Table 3 presents estimated changes in milk consumption for various price increases. For simplicity, all cow's milk products receive the same nominal price change, as do all non-dairy milk products.

Assuming carbon footprints of 1.8 kg CO₂e per kilogram of cow's milk and 0.6 kg CO₂e per kilogram of soy milk, the first three rows correspond to carbon prices of \$12, \$30, and \$90 per metric ton of CO₂, respectively. The final row corresponds to an extreme counterfactual with a large price increase for cow's milk and no price change for soy milk.

Comparing the changes for the logit and random coefficients specifications, there is less substitution from cow's milk to soy milk after accounting for heterogeneous preferences.

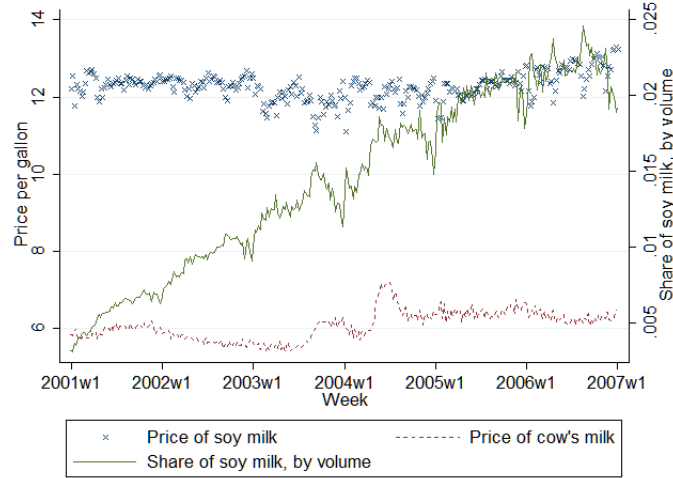


FIGURE 3.2. Prices and market share growth, national averages.

By volume, soy milk accounts for only 2.5% of sales (by volume) in the data set, so the percentage changes in soy milk consumption are relatively small when translated to volumes. In the extreme counterfactual, soy milk's share of milk consumption rises to 3.7%.

On the other hand, the change in the *level* of milk consumption may be more substantial, considering that dairy industry is responsible for an estimated 4% of anthropogenic greenhouse gas emissions (FAO, 2010). However, since I have only modeled demand for milk, I cannot say how reduced milk consumption will affect consumption of other goods, and the net effect on emissions depends on these other consumption changes, too. In other words, the model does not tell us what the outside option's carbon footprint is.⁶

3.4. Milk consumption over time

The analysis in this section is based on a larger data set on retail consumption, covering in 48 regional markets within the US between 2001 and 2006.⁷

Soy milk's national share (by volume) of retail milk sales grew from about 0.5% in 2001 to over 2% in 2006, increasing in every regional market in the US. There were no dramatic price changes, and the cross-price elasticities are low anyways, so price dynamics cannot possibly explain the time trend.

⁶It could be argued that, unless expenditure on milk is replaced by expenditure on products like beef, which has a very high carbon footprint, it's likely that decreased milk consumption will be emissions-saving. At \$3/gallon and 1.8 kg CO₂e per kg of milk, cow's milk purchases correspond to about 2.3kg CO₂e per dollar spent. In contrast, US emissions in 2009 were around 6.6 Gt CO₂e (EPA, 2011), or about 0.47 2.3kg CO₂e per dollar of GDP.

⁷I do not use the larger data set to estimate the random coefficients model because it does not include unit cost instruments and there is no location information beyond metropolitan area, prohibiting association of stores with zip-code level demographic information. Unfortunately, stores cannot be linked across years, making estimation of a dynamic model difficult.

This trend is in striking contrast to the results of section 3.3 – given my estimates, even a very high \$1/gallon tax on cow’s milk would affect soy milk’s share of milk sales by less than the five-year change, which occurred without any significant price changes.

In other words, my estimates suggest that the effect of forcing consumers to internalize the externalities of GHG emissions embodied in milk would have a smaller effect than whatever drove the trend displayed in Figure 2. This is a strong suggestion that there is more to food consumption patterns than static relationships between prices and quantities.

Many explanations for the trend are possible – for example, information diffusion and consumption externalities. One might attribute some of the trend to growth in the soy milk product set, but this merely shifts the burden of explanation: why did the set of soy milk products grow over this period?

Another possibility is that consumers gradually discovered their soy milk preferences over time through experimentation. Tables 4 and 5 present evidence corroborating this explanation. In 2001, in-store soy milk promotions are associated with significantly elevated soy milk consumption in subsequent weeks. In 2006, there is no such lagged promotional effect – in fact, the small promotions in 2006 are associated with marginally significant lagged decreases in soy milk consumption, but these changes are far smaller in magnitude than the effects in 2001.

Thus, promotions in 2001 may have encouraged consumers to experiment with soy milk in 2001, resulting in higher consumption in weeks following the promotion due to consumers who discovered that they enjoyed soy milk. However, by 2006, most consumers might have been informed about their preferences, causing the lagged promotional effect to disappear. The slight dip following promotions in 2006 could be due to stocking up during promotional periods.

While it is hard to rationalize this pattern with other explanations, the presence of learning-by-consuming is compatible with other factors influencing the soy milk’s growth. In fact, it may be very difficult to fully explain such a long period of growth simply with a model of rational experimentation with a new product (i.e., why didn’t everybody experiment within the first year?). Social aspects to information diffusion (perhaps many people never seriously consider trying soy milk until hearing that a neighbor has switched to using it on her breakfast cereal) could explain prolonged adoption, as could consumption externalities (learning about appropriate consumption of a product from one’s neighbors). Determining whether these or other sources of dynamics are important in driving changes in food consumption is an important topic for future work.

References

Berry, Steven, James Levinsohn, and Ariel Pakes, “Automobile Prices in Market Equilibrium,”

- Econometrica*, July 1995, 63 (4), 841–90.
- Berry, Steven T.**, “Estimating Discrete-Choice Models of Product Differentiation,” *RAND Journal of Economics*, 1994, 25 (2), 242–262.
- Carlsson-Kanyama, Annika and Alejandro D. Gonzalez**, “Potential contributions of food consumption patterns to climate change,” *The American Journal of Clinical Nutrition*, 2009, 89.
- EPA**, “U.S. Greenhouse Gas Inventory Report,” 2011. United States Environmental Protection Agency report.
- Eshel, Gidon and Pamela A. Martin**, “Diet, Energy, and Global Warming,” *Earth Interactions*, 2006.
- FAO**, “Greenhouse Gas Emissions from the Dairy Sector,” 2010. Food and Agriculture Organization report.
- Pelletier, Nathan**, “Environmental performance in the US broiler poultry sector: Life cycle energy use and greenhouse gas, ozone depleting, acidifying and eutrophying emissions,” *Agricultural Systems*, 2008, 98 (2), 67–73.
- , **Rich Pirog, and Rebecca Rasmussen**, “Comparative life cycle environmental impacts of three beef production strategies in the Upper Midwestern United States,” *Agricultural Systems*, 2010, 103 (6), 380–389.
- Thoma, Greg, Jennie Popp, Darin Nutter, David Shonnard, Richard Ulrich, Marty Matlock, Dae Soo Kim, Zara Neiderman, Nathan Kemper, Cashion East, and Felix Adom**, “Regional Analysis of Greenhouse Gas Emissions from Milk Production Practices in the United States,” 2010. Innovation Center for US Dairy report.
- Train, Kenneth**, *Discrete Choice Methods with Simulation*, Cambridge UP, 2009.

Table 4: Effect of promotions, 2001

	$\log(Q_{soy})$		
$\log(P_{soy})$	-1.426*** (0.0181)	-1.192*** (0.0192)	-1.428*** (0.0180)
$\log(P_{cow})$	0.0130 (0.0141)	0.0187 (0.0152)	0.0136 (0.0141)
small promo	-0.00898** (0.00367)	0.0324*** (0.00391)	-0.00989*** (0.00367)
big promo	-0.0471*** (0.00588)	0.0404*** (0.00624)	-0.0482*** (0.00587)
recent promo, small	0.0175*** (0.00337)	0.0258*** (0.00362)	0.0164*** (0.00337)
recent promo, big	0.0390*** (0.00806)	0.0875*** (0.00864)	0.0370*** (0.00805)
new product		0.0167* (0.00889)	-0.0473*** (0.00829)
recent new product		0.0513*** (0.00409)	-0.0333*** (0.00390)
# of soymilk brands	0.0716*** (0.000754)		0.0735*** (0.000777)
Observations	58,547	58,547	58,547
R-squared	0.332	0.229	0.334
Standard errors in parentheses			
*** p<0.01, ** p<0.05, * p<0.1			
All regressions include store fixed effects and quadratic time trends.			

Q_{soy} is the total volume of soy milk sales in a given store-week, summing over products. P_{cow} and P_{soy} refer to total revenues across products divided by total volume.

“Small promo” refers to a dummy variable which is equal to one when between one and three soy milk products are on sale. “Big promo” is equal to one when four or more soy milk products are on sale.

The “new product” dummy variable is equal to one in store-weeks where a product appears in a store for the first time. Stores cannot be matched across years, so observations in January are excluded from the regressions.

The “recent promo” variables are equal to one when there is no sale in the current week, but the store had a promotion of the appropriate size in one of the four previous weeks. The “recent new product” variable is coded similarly.

Table 5: Effect of promotions, 2006

	$\log(Q_{soy})$		
$\log(P_{soy})$	-1.284*** (0.0115)	-1.173*** (0.0121)	-1.284*** (0.0115)
$\log(P_{cow})$	0.0143 (0.00978)	0.00615 (0.0104)	0.0142 (0.00978)
small promo	-0.0115*** (0.00271)	0.00573** (0.00287)	-0.0115*** (0.00271)
big promo	-0.00673** (0.00292)	0.0272*** (0.00308)	-0.00665** (0.00292)
recent promo, small	-0.00561** (0.00283)	0.00198 (0.00301)	-0.00556** (0.00283)
recent promo, big	-0.00560* (0.00299)	0.00208 (0.00318)	-0.00569* (0.00299)
new product		0.0311*** (0.00911)	-0.0162* (0.00859)
recent new product		0.0406*** (0.00416)	-0.0110*** (0.00395)
# of soymilk brands	0.0389*** (0.000417)		0.0391*** (0.000422)
Observations	66,983	66,983	66,983
R-squared	0.305	0.214	0.305

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

All regressions include store fixed effects and quadratic time trends.