

Dynamic Discrete Choice Estimation of Agricultural Land Use

Paul T. Scott *

Toulouse School of Economics

December 15, 2013

Abstract

I develop a new framework for analyzing land use change with dynamically optimizing landowners. My empirical approach allows for unobservable heterogeneity and avoids the burden of explicitly modeling the evolution of market-level state variables like input and output prices. Using a rich new data set on land use in the United States, I estimate a relatively large long-run cropland-price elasticity of 0.3. Compared to static estimates using the same data, my dynamic estimates suggest that biofuels production leads to dramatically more land use change and substantially smaller price increases in the long run.

Keywords: agricultural supply estimation, dynamic discrete choice, land use change, biofuels policy

*I am grateful to Jan De Loecker, Bo Honoré, Steve Berry, and Myrto Kalouptsidi for all their support. I thank Lanier Benkard, Sylvain Chassang, Joseph Hotz, Robert Miller, Eduardo Morales, Andriy Norets, Sergio Pastorello, Mar Reguant, Nathan Yang, and various seminar participants for helpful discussions and suggestions. I thank Wolfram Schlenker and Michael Roberts for generously providing the weather data used in this paper.

21 allée de Brienne, 31015 Toulouse Cedex 6, France. Email: ptscott@gmail.com.

1 Introduction

Empirical studies of land use change typically treat landowners as static decision makers despite the fact that 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., crop rotation). In this paper, I formulate and implement a flexible and tractable empirical approach for analyzing land use based on a model of dynamically optimizing landowners.

The effects of many controversial policies concerning greenhouse gas mitigation, ecological destruction, and agricultural policy depend crucially on how land use patterns respond to economic changes. 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 central 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. Thus, biofuels mandates effectively increase crop demand.² In the US, a staggering 35-40% of corn production has been used to produce ethanol in recent years (US EPA, 2011). 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 elasticities (Roberts and Schlenker, 2013) and in simulating equilibrium responses (Searchinger et al., 2008; Tyner et al., 2010).

Static models remain common in empirical work on land use (Chomitz and Gray, 1996; Fezzi and Bateman, 2011; Souza-Rodrigues, 2012), and some studies incorporate state dependence without forward-looking dynamics (Nerlove, 1956; Lubowski et al., 2006). However, static and myopic models of land use are likely to understate long-run land use responses. Intuitively, landowners may respond more to long-run changes in the process governing price

¹ In 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.

² There 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.

changes than they respond to year-to-year price variation in the data; i.e., landowners may be more willing to pay the costs associated with bringing new land into crops in response to a long-run price increase than in response to a temporary price increase. This creates an external validity problem when static or myopic models are used for counterfactual policy analysis; e.g., we might only observe short-run variation in the data, but one would expect policy like the biofuels mandate to have a long-run price impact. Unfortunately, existing studies on land use which account for forward-looking behavior are rare and largely confined to models of irreversible decisions (Irwin and Bockstael, 2002; Vance and Geoghegan, 2002).

Using a new empirical framework, I estimate a dynamic discrete choice model of cropland use in the United States with forward looking landowners, finding a long-run elasticity of crop acreage with respect to crop prices in the neighborhood of 0.3. This elasticity is roughly ten 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. Comparing my long-run elasticity estimates using dynamic model to static elasticities based on the same data, I find that taking dynamics into account leads to a 160% larger land use effect and a 78% smaller price increase in the long run.

This paper’s main methodological contribution is to show how a dynamic model of land use can be estimated using a linear regression equation and without modeling how market-level state variables evolve. The approach is analogous to the Euler equation approach which has a long history in the context of single-agent dynamic models with continuous choice variables (Hall, 1978). While most dynamic discrete choice estimation approaches rely on evaluating Bellman equations (Rust, 1987; Aguirregabiria and Mira, 2002) or simulating the model (Hotz et al., 1994; Bajari et al., 2007), my approach relies on estimating the realized path of continuation values. Given that agents are small and have rational expectations, valid moments for estimation can be constructed using realized continuation values despite the fact that agents make decisions based on earlier expectations of the continuation values.³

Avoiding the need to model the evolution of market-level state variables is a major advantage in the context of agricultural land use. The set of market-level variables which influence farmers’ expected returns is large (e.g., input and output prices, technological conditions, government policies, and crop stocks), and dealing with such a large state space would make other estimation strategies infeasible without strong simplifying assumptions. Furthermore, I can allow for unobservable supply shocks which may be serially correlated – because they are unobservable, such shocks are difficult to handle when the empirical approach requires a model of how all state variables evolve. The inclusion of these error terms is appealing given

³ Similar regression equation constructions could be applied to other single-agent 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 that prices evolve according to a particular sort of process (Hendel and Nevo, 2006; Gowrisankaran and Rysman, 2011).

the infeasibility of perfectly measuring every variable which shifts farmers' incentives.⁴

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 choice probabilities.

The main data set is a rich new panel of land cover data spanning the entire contiguous United States in recent years. I also construct a measure of expected returns to cropland 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. However, cross-sectional variation in expected yields results in some cross-sectional variation in my measure of expected returns.

In Section 2, I lay out a binary choice model of land use and derive a regression equation. In Section 3, I describe data sources, the construction of the land use panel data set, and the measurement of expected returns. Details regarding estimation are treated in Section 4, including the extension to unobservable heterogeneity. Section 5 presents the results; Section 6 considers implications for biofuels policy; and Section 7 concludes.

2 Empirical framework

This section presents 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 4, I explain how I estimate the model with different observable types and with a mixture of unobservable types.

⁴ Typically, the only error terms empirical models of dynamic discrete processes allow for are conditionally independent, ruling out serially correlated errors. However, it is hard to imagine measurement errors ever satisfying such restrictive assumptions in the context of crop agriculture. Given the limited availability of local price data, differences between measured and actual prices may be nontrivial, and there is little reason to doubt that such errors will be serially correlated.

⁵ For 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.

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 $\mathbf{J} = \{crops, other\}$.⁶ Let j_{it} denote the land use of field i in year t .

There are two types of state variables in the model. First, the *field state*, k_{it} , represents characteristics specific 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 information necessary to determine expected returns in the current period for each 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

$$\pi(j, k, \omega_t, \nu_{it}) = \alpha_0(j, k) + \alpha_R R_j(\omega_t) + \xi_{jk}(\omega_t) + \nu_{jit} \quad (1)$$

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 $\alpha = (\alpha_R, \{\alpha_0(j, k)\}_{j \in \mathbf{J}, k \in \mathbf{K}})$ 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 3 explains how I construct the expected returns variable.

The inclusion of the unobservable shock ξ in the profit equation is important given the limitations of available data on farmer's expected returns. Local data of input and output prices for crops are not available, and there is very limited data on the returns to non-crop land uses such as pasture land.

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 difficult to identify the effect of the field state on both intercepts and productivity in a short panel. Furthermore, it is relatively common in the literature to restrict dynamics to only affect an intercept term

⁶The empirical approach I describe can be generalized to larger discrete choice sets. See Appendix A.6.

(Claassen and Tegene, 1999; Munroe et al., 2004; Wu et al., 2004), so I adopt this strategy as a starting point.

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 that landowners are price-taking agents is plausible.⁷ While the assumption that there are no externalities across fields is arguably less plausible as a general claim, few studies have made any attempt to model the economies of space in crop agriculture.⁸ This paper focuses on developing a model of land use which can account for dynamic decision making, and treating dynamics and spatial effects together may be an important topic for future work.

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 separately 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.⁹ 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

⁷The 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.

⁸Anecdotal 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.

⁹A 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.

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, planting crops may be more profitable 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

$$\kappa^+(j, k) = \begin{cases} 0 & \text{if } j = \text{crops} \\ \min\{k + 1, \bar{k}\} & \text{if } j = \text{other}. \end{cases} \quad (2)$$

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 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 $F(\nu_{jit}) = \exp(-\exp(-\nu_{jit}))$.¹⁰

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:

$$V(k_{it}, \omega_t, \nu_{it}) \equiv \max_{j^*} E \left(\sum_{s \geq t} \beta^{s-t} \pi(j^*(k_{is}, \omega_s, \nu_{is}), k_{is}, \omega_s, \nu_{is}) | k_{it}, \omega_t, \nu_{it} \right). \quad (3)$$

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

My empirical approach is very flexible with respect to the process governing the evolution of the market state ω_t . The process must be well behaved enough for the value function to exist,¹¹ 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

¹⁰The 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.

¹¹See Bhattacharya and Majumdar (1989) for regularity conditions on G which guarantee the existence of the value function.

be modeled explicitly or estimated. Hereafter, I will mainly use t subscripts on functions and variables which depend on ω_t ; e.g., $V_t(k_{it}, \nu_{it}) \equiv V(k_{it}, \omega_t, \nu_{it})$, and $R_{jt} \equiv R_j(\omega_t)$.

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

$$\bar{V}_t(k) \equiv \int \dots \int V_t(k, (\nu_1, \dots, \nu_J)) dF(\nu_1) \dots dF(\nu_J). \quad (4)$$

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

$$v_t(j, k) \equiv \bar{\pi}_t(j, k) + \beta E_t \left[\bar{V}_{t+1} \left(\kappa^+(j, k) \right) \right]. \quad (5)$$

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 $p_t(j, k) \equiv \Pr(j_{it} = j | k_{it} = k, \omega_t)$,

$$p_t(j, k) = \frac{\exp(v_t(j, k))}{\sum_{j' \in J} \exp(v_t(j', k))}. \quad (6)$$

Assumption 2 also implies a convenient expression for the mean value function:

$$\bar{V}_t(k) \equiv \ln \left(\sum_{j \in J} \exp(v_t(j, k)) \right) + \gamma \quad (7)$$

where γ is Euler's gamma.

2.2 Deriving a regression equation

In this section, I derive a linear regression equation for the model presented above. Appendix A.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 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 (6) across j :

$$\ln \left(\frac{p_t(j, k)}{p_t(j', k)} \right) = v_t(j, k) - v_t(j', k). \quad (8)$$

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

$$\bar{\pi}_t(j, k) - \bar{\pi}_t(j', k) - \ln \left(\frac{p_t(j, k)}{p_t(j', k)} \right) = \beta E_t \left[\bar{V}_{t+1}(\kappa^+(j', k)) \right] - \beta E_t \left[\bar{V}_{t+1}(\kappa^+(j, k)) \right] \quad (9)$$

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 (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:

$$\begin{aligned} \bar{\pi}_t(j, k) - \bar{\pi}_t(j', k) - \ln \left(\frac{p_t(j, k)}{p_t(j', k)} \right) &= \beta \left(\bar{V}_{t+1}(\kappa^+(j', k)) - \bar{V}_{t+1}(\kappa^+(j, k)) \right) \\ &\quad + \varepsilon_t^V(j', k) - \varepsilon_t^V(j, k) \end{aligned} \quad (10)$$

where

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

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 $v_t(j, k)$ from equation (7), and substituting using equation (6):

$$\forall j : \bar{V}_t(k) = v_t(j, k) - \ln(p_t(j, k)) + \gamma. \quad (11)$$

Equation (11) is a special case of Lemma 1 in Arcidiacono and Miller (2011), and versions of it also appear in Altug and Miller (1998) and Arcidiacono and Ellickson (2011).

Note well that equation (11) holds for any land use j . It is particularly convenient to

apply equation (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 $t + 1$.¹²

Replacing the continuation values in equation (10) using equation (11),

$$\begin{aligned} \bar{\pi}_t(j, k) - \bar{\pi}_t(j', k) - \ln\left(\frac{p_t(j, k)}{p_t(j', k)}\right) &= \beta(v_{t+1}(j_{re}, \kappa^+(j', k)) - v_{t+1}(j_{re}, \kappa^+(j, k))) \\ &\quad - \beta\left(\ln\left(\frac{p_{t+1}(j_{re}, \kappa^+(j', k))}{p_{t+1}(j_{re}, \kappa^+(j, k))}\right)\right) \\ &\quad + \varepsilon_t^V(j', k) - \varepsilon_t^V(j, k). \end{aligned} \quad (12)$$

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

$$v_{t+1}(j_{re}, \kappa^+(j', k)) - v_{t+1}(j_{re}, \kappa^+(j, k)) = \bar{\pi}_{t+1}(j_{re}, \kappa^+(j', k)) - \bar{\pi}_{t+1}(j_{re}, \kappa^+(j, k)) \quad (13)$$

Finally, the Euler equation comes from substituting equation (13) into equation (12):

$$\begin{aligned} \bar{\pi}_t(j, k) - \bar{\pi}_t(j', k) - \ln\left(\frac{p_t(j, k)}{p_t(j', k)}\right) &= \beta(\bar{\pi}_{t+1}(j_{re}, \kappa^+(j', k)) - \bar{\pi}_{t+1}(j_{re}, \kappa^+(j, k))) \\ &\quad - \beta\ln\left(\frac{p_{t+1}(j_{re}, \kappa^+(j', k))}{p_{t+1}(j_{re}, \kappa^+(j, k))}\right) \\ &\quad + \varepsilon_t^V(j', k) - \varepsilon_t^V(j, k). \end{aligned} \quad (14)$$

As explained above, the left-hand side represents the difference in continuation 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, plus a term which corrects for the fact that this action isn't always optimal. This correction is possible thanks to equation (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 A.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 (14) can be rearranged into the following linear regression equation:

$$Y_{tk} = \tilde{\Delta}\alpha_{0k} + \alpha_R\Delta R_t + \tilde{\Delta}\xi_{tk} + \Delta\varepsilon_{tk}^V \quad (15)$$

¹²Renewal actions are a special case of *finite dependence*, defined by Arcidiacono and Miller (2011).

where

$$\begin{aligned}
Y_{tk} &\equiv \ln \left(\frac{p_t(crops, k)}{p_t(other, k)} \right) + \beta \ln \left(\frac{p_{t+1}(crops, 0)}{p_{t+1}(crops, \kappa^+(other, k))} \right) \\
\tilde{\Delta}\alpha_{0k} &\equiv \alpha_0(crops, k) - \alpha_0(other, k) \\
&\quad + \beta (\alpha_0(crops, 0) - \alpha_0(crops, \kappa^+(other, k))) \\
\Delta R_t &\equiv R_{crops, t} - R_{other, t} \\
\tilde{\Delta}\xi_{tk} &\equiv \xi_{crops, k, t} - \xi_{other, k, t} + \beta (\xi_{crops, 0, t+1} - \xi_{crops, \kappa^+(other, k), t+1}), \\
\Delta\varepsilon_{tk}^V &\equiv \varepsilon_t^V(crops, k) - \varepsilon_t^V(other, k).
\end{aligned} \tag{16}$$

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 constructed dependent variable.

The importance of the perfectly competitive setting is that the econometrician effectively observes the relevant counterfactuals in which agents had taken different actions than they actually took. That is, suppose that field i is in land use j_{it} in year t , resulting in field state $k_{i,t+1} = k$ in year $t + 1$. The realized choice probabilities in period $t + 1$ for fields in state $k' \neq k$ are a valid estimate of what the choice probabilities would have been for field i in period $t + 1$ if field i had taken a different course of actions leading to $k_{i,t+1} = k'$. In an imperfectly competitive setting, we don't observe these counterfactuals because the agent's own decision in period t has a non-trivial impact on the state of the world in period $t + 1$.

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.¹³ 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 can 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

¹³To 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.

Thesmar, 2002). The fact that the model is not identified in its full generality is easy to see in this setting, for estimating equation (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 estimates of equation (15). 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.¹⁴

3 Data and measurement

In this section, I describe the two data inputs used to estimate a model of US cropland. First, I describe how I construct a rich panel data set on land use in the United States. Subsequently, I explain the several steps involved in constructing a measure of expected returns R .

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.¹⁵ 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 production accounts for roughly 23% of worldwide output in caloric terms (Roberts and Schlenker, 2013).

3.1 Land use data

The National Agricultural Statistics Service's Cropland Data Layer (CDL) features finely detailed land cover data for the United States.¹⁶ The CDL's land cover classifications include over 50 different crops as well as about 20 non-crop classifications (e.g., grassland, forest,

¹⁴Specifically, $\alpha_0(\text{crops}, k)$ can be recovered from $\tilde{\Delta}\alpha_0(k)$ 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., $\alpha_0(\text{crops}, 1) = \beta^{-1} \left((1 + \beta) \alpha_0(\text{crops}, 0) - \tilde{\Delta}\alpha_0(0) \right)$.

¹⁵Based on export values, FAOSTAT Database on Agriculture, visited 10/31/2012.

¹⁶See <http://nassgeodata.gmu.edu/CropScape/> for data and Boryan et al. (2011) for a description data inputs, classification system, and validations procedures.

water, several levels of developed land). Since 2008, the annual CDL data have provided field-level resolution (30m or 56m) of the entire contiguous United States.

I construct a panel of land use outcomes using the CDL data for 2006-2012, avoiding CDL data from before 2006 because they are less reliable and cover fewer states. 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.

After excluding water, protected land, and developed land, remaining points were classified as cropland or non-cropland. Table 1 lists the share of points in cropland by point and the the initial year in the panel by state.

The CDL data performs poorly when it comes to distinguishing among unmanaged grassland, pasture, and hay.¹⁷ While hay might be more naturally categorized as cropland, this limitation of the data leads me to assign it to the non-cropland category, lumping it together with grassland, pasture and various forms of unmanaged land such as shrubland, forests, and wetlands.

Further details regarding the land use data are included in Appendix B.1.

3.2 Crop-specific returns

Before constructing a measure of the average returns to cropland, I construct 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 **C**). These crops account for about 94% of harvested cropland (excluding hay) in the US, according to the 2007 Census of Agriculture.

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 also 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

$$R_{czt} = (P_{czt} - e_{czt}) \cdot YIELD_{czt} \quad (17)$$

where P_{czt} represents the expected output price (including government payments), $YIELD_{czt}$

¹⁷The 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?

Table 1: Summary statistics for land use panel

State	counties in sample	points in sample	share in crops	initial year
Alabama	49	121006	0.050	2009
Arkansas	26	58068	0.541	2007
California	15	74105	0.260	2008
Colorado	30	187300	0.135	2009
Georgia	67	88362	0.171	2009
Idaho	38	202990	0.085	2008
Illinois	102	169048	0.726	2007
Indiana	92	108929	0.570	2007
Iowa	99	180652	0.707	2007
Kansas	105	277396	0.387	2007
Kentucky	60	68197	0.194	2009
Louisiana	32	70728	0.207	2007
Maryland	22	25276	0.236	2009
Michigan	64	122998	0.264	2008
Minnesota	78	200920	0.495	2007
Mississippi	49	89742	0.221	2007
Missouri	81	152978	0.301	2007
Montana	49	443806	0.078	2008
Nebraska	90	237878	0.359	2007
New Jersey	13	12372	0.110	2009
New York	49	138677	0.080	2009
North Carolina	77	126148	0.160	2009
North Dakota	53	224953	0.409	2007
Ohio	72	102982	0.460	2007
Oklahoma	60	186974	0.216	2007
Oregon	20	181990	0.046	2008
Pennsylvania	56	118894	0.107	2009
South Carolina	33	74199	0.097	2009
South Dakota	62	236735	0.317	2007
Tennessee	55	81522	0.161	2009
Texas	174	547901	0.178	2009
Utah	13	140410	0.012	2009
Washington	14	74520	0.255	2007
Wisconsin	65	156457	0.223	2007
Wyoming	11	113773	0.022	2009

is the expected yield (output per acre), and e_{czt} represents a unit production cost.¹⁸

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 is arguably reasonable to use realized costs 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 expected yields are constructed. Further details can be found in Appendix B.

3.3 Expected yields

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

$$\ln(YIELD_{czt}) = \theta_{cz} + \theta_{cw}W_{zt} + \theta_c t + \varepsilon_{czt}^Y \quad (18)$$

where W_{zt} includes weather variables described in Appendix B, θ_{cz} is a county-level fixed effect, and θ_c 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 (18) using the county's historical mean values of W_{zt} .

For some county-crop pairs, NASS yield data is 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 is included in the fixed effects regression (see Appendix B.3 for details).

Figures 6-8 illustrate that my expected yields variables are good forecasts of average county-level yields between 2006 and 2012 (noting that no data from 2006-2012 were used to make the forecasts). The yield forecasts prove effective even for counties with fixed effects imputed from neighbors.¹⁹

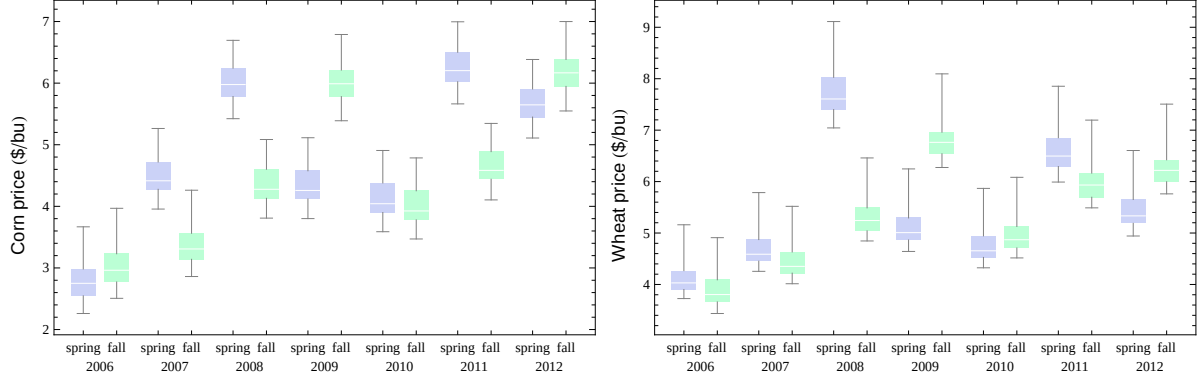
3.4 Expected prices

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

¹⁸Strictly speaking, equation (17) 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.

¹⁹For 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 6-8 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: 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 season.

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 1997-2012 data:

$$P_{crt}^{rec} = \theta_{1cr} + \theta_{2cr}P_{ct}^{fut} + \varepsilon_{crt}^{P} \quad (19)$$

where P_{crt}^{rec} is the price received for crop c in state r in year t . 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 4.

Expected market prices are defined as fitted values to equation (19):

$$\tilde{P}_{crt} \equiv \hat{\theta}_{1cr} + \hat{\theta}_{2cr}P_{ct}^{fut}. \quad (20)$$

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 perhaps precede it slightly, since crop insurance sign ups, fertilizer application, and some input purchases must be made in advance of planting). 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 (19), corresponding to the two different planting seasons. Expected market-level prices for each county are taken from one version or the other, depending on which crops are typically planted in the county (see Appendix B for details).

Figure 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.²⁰ The expected price received by farmers is

$$P_{czt} \equiv \max \left\{ \tilde{P}_{czt}, TP_{ct} \right\},$$

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

3.5 Aggregating over crops

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

$$R_{crops,z,t} = \frac{\sum_{c \in \mathbf{C}_z} A_{cr} R_{czt}}{\sum_{c \in \mathbf{C}_z} A_{cr}} \quad (21)$$

where s denotes the US state containing county z , A_{cr} is the average annual harvested area of crop c in state r (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 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

$$\frac{\sum_{c \in \mathbf{C}_z} A_{czt}}{\sum_{c \in \mathbf{C}} A_{czt}} > .75 \quad (22)$$

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 B.5 for details).

After all exclusions, there are 1975 counties remaining in my sample. Collectively, they contain over 90% of US cropland, according to the 2007 Census of Agriculture. Table 1 displays the number of counties in the sample by US state, and Figure 4 is a map illustrating which counties are in the sample.

²⁰In 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.

3.6 Net expected returns

Estimation is based on the difference in crop returns between cropland and non-cropland:

$$\Delta R_{z,t} = R_{crops,z,t} - R_{other,z,t}.$$

For the returns to non-cropland, I use pasture land rental rates published on line by NASS. County-year pasture land rental rates are used when available, and state-year rental rates otherwise. When neither county- nor state-level rental rates are available, I simply set $R_{other,z,t} = 0$.

Pasture land rental rates generally have lower variance than cropland returns, and incorporating them into the measure of $\Delta R_{z,t}$ makes little difference to the results, for cropland returns account for most of the variation in $\Delta R_{z,t}$. This is fortunate because information on the returns to pasture land is much harder to come by than information on the returns to cropland. While rental rates are naturally correlated with returns, they may be imperfectly correlated, and they may respond little to short-run variation in returns.

4 Estimation

The model presented in Section 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 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.²¹ In Section 4.2, I extend the estimation of the model to observable heterogeneity.

4.1 Choice probability smoothing

In principle, conditional choice probability estimates can 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 in Section 2.

I smooth choice probability estimates by taking a weighted average of frequency estimates, with weights inversely proportional to distances between counties. The smoothed estimates

²¹Initially, 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.

are:

$$\hat{p}_{zt}(crops, k) = \frac{\sum_{z' \in Z_r} w_{zz'} \sum_{i \in I_{z'}} D_{it}(j, k)}{\sum_{z' \in Z_r} w_{zz'} \sum_{i \in I_{z'}} D_{it}(k)} \quad (23)$$

where $D_{it}(j, k) = 1$ if $j_{it} = j$ and $k_{it} = k$ and $D_{it}(j, k) = 0$ otherwise, $D_{it}(k)$ is defined similarly, I_z is the set of fields in county z , and Z_r is the set of counties within US state r . The smoothing weight $w_{zz'}$ is inversely proportional to the square of the distance between counties: $w_{zz'} \equiv (1 + d_{zz'})^{-2}$, where $d_{zz'}$ is the distance between z and z' measured in kilometers. If z and z' are not within the same US state, then $w_{zz'} = 0$, even if they are adjacent.²² Cross-county smoothing weights are small; the median county has 98% of the total weight on its own frequency estimate, and the lowest own-county weight is about 95%.

After smoothing, I use the resulting conditional choice probabilities to construct dependent variables Y_{zt} , as in the derivation of equation (15).

4.2 Extension to unobservable heterogeneity

Ignoring unobservable heterogeneity can 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 lead to exaggerated switching cost estimates. 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 and estimating 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.

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.

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

²²I 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.

can be explained by field states may be evidence of unobservable heterogeneity. Put broadly, identification of unobserved heterogeneity comes from correlations in agents' decisions over time which are cannot be explained by state dependence. Consequently, the identification of unobservable heterogeneity relies crucially on the model of how field states evolve.²³

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

$$\mu_z(\zeta, k) \equiv Pr(k_{i1} = k, \zeta_i = \zeta | i \in I_z) \quad (24)$$

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

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:

$$q_{i\zeta} \equiv Pr(\zeta_i = \zeta | \mathbf{j}_i, \mathbf{k}_i) = \mu_z(\zeta, k_{i1}) \prod_{t=1}^T p_{z\zeta t}(j_{it}, k_{it}) \quad (25)$$

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 $\hat{p}_{z\zeta t}^{(m)}(crops, k)$ and prior probabilities $\hat{\mu}_{z\zeta}^{(m)}(k)$:

$$q_{i\zeta}^{(m)} = \hat{\mu}_{z\zeta}^{(m)}(k_{i1}) \prod_{t=1}^T \hat{p}_{z\zeta t}^{(m)}(j_{it}, k_{it}). \quad (26)$$

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

$$\hat{p}_{z\zeta t}^{(freq, m)}(j, k) = \frac{\sum_{i \in I_z} q_{i\zeta}^{(m-1)} D_{it}(j, k)}{\sum_{i \in I_z} q_{i\zeta}^{(m-1)} D_{it}(k)}, \quad (27)$$

$$\hat{p}_{z\zeta t}^{(m)}(j, k) = \frac{\sum_{z' \in Z_s} w_{zz'} \hat{p}_{z'\zeta t}^{(freq, m)}(j, k)}{\sum_{z' \in Z_s} w_{zz'}}, \quad (28)$$

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, $\hat{p}_{z\zeta t}^{(m)}(other, k) = 1 - \hat{p}_{z\zeta t}^{(m)}(crops, k)$.

²³For 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).

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

$$\hat{\mu}_{z\zeta}^{(m)}(k) = \frac{\sum_{i \in I_z} q_{i\zeta}^{(m-1)} D_{i1}(k)}{\sum_{i \in I_z} q_{i\zeta}^{(m-1)}}. \quad (29)$$

The EM algorithm involves iterating on the E step (equation (26)) and the M step (equations (27-29)) until convergence.

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

$$\sum_z \sum_{i \in I_z} \log \left(\sum_{\zeta} \mu_z(\zeta, k_{i1}) \prod_{t=1}^T p_{z\zeta t}(j_{it}, k_{it}) \right). \quad (30)$$

If it were the case that only own-county weights were non-zero, the above algorithm would indeed be a traditional implementation of the EM algorithm to maximize a likelihood function. 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 (27-28) 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 (26-29). 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.

4.3 Identifying assumptions

The returns to competing land uses are plausibly correlated with variation in crop returns both in the cross section and over time. Note that my only measure of returns to non-cropland is an estimate of rental rates for pasture land. For some counties, I am missing these data, and it is a crude measure in any case. Any unmeasured variation in returns to the non-cropland alternative will be absorbed by the unobservable shock ξ . Livestock and crop output prices are correlated because feed grains are an important input in livestock production. Therefore, the expected returns to grazing livestock (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 allow for some 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.

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 :

$$Y_{nt} = \tilde{\Delta}\alpha_{0n} + \alpha_{R\zeta}\Delta R_{nt} + \tilde{\Delta}\xi_{nt} + \Delta\varepsilon_{nt}^V \quad (31)$$

where the differenced variables are defined in equation (16).

Equation (31) seems to call for a standard fixed effects estimation strategy. However, this would implicitly require implausible identifying assumptions.

The rational expectations assumption implies the following moments:

$$\forall t : E \left[\Delta\varepsilon_{nt}^V \Delta R_{nt} \right] = 0. \quad (32)$$

However, fixed effects estimation requires a stronger assumption:

$$\forall t, t' : E \left[\Delta\varepsilon_{nt}^V \Delta R_{nt'} \right], \quad (33)$$

which is not implied by the model and indeed unlikely to be true when considered carefully. For example, condition (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 (31),

$$Y_{n,t+1} - Y_{nt} = \alpha_{R\zeta}(\Delta R_{n,t+1} - \Delta R_{nt}) + (\tilde{\Delta}\xi_{n,t+1} - \tilde{\Delta}\xi_{nt}) + (\Delta\varepsilon_{n,t+1}^V - \Delta\varepsilon_{nt}^V). \quad (34)$$

We can then consistently estimate $\alpha_{R\zeta}$ using equation (34) with ΔR_{nt} as an instrument for $(\Delta R_{n,t+1} - \Delta R_{nt})$ given the identifying assumption:

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

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

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

constant term. I use two-stage GMM, first using two-stage least squares, and then using the estimated residuals to construct a weighting matrix for a second GMM estimation.

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:²⁴

$$\ln \left(\frac{p_{zt}(\text{crops}, k)}{p_{zt}(\text{other}, k)} \right) = \alpha_{0zk} + \alpha_R \Delta R_{zt} + \xi_{ztk}. \quad (35)$$

Equation (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

$$Y_{tk} = \ln \left(\frac{p_{zt}(\text{crops}, k)}{p_{zt}(\text{other}, k)} \right) + \beta \ln \left(\frac{p_{z,t+1}(\text{crops}, 0)}{p_{z,t+1}(\text{crops}, \kappa^+(\text{other}, k))} \right). \quad (36)$$

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

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 (35), but with no dependence on k :

$$\ln \left(\frac{p_{zt}(\text{crops})}{p_{zt}(\text{other})} \right) = \alpha_{0z} + \alpha_R R_{zt} + \xi_{zt}. \quad (37)$$

4.5 Defining elasticities

Elasticities for static models (without unobservable heterogeneity) are computed as follows:

$$\left(\sum_z A_{zt} \right)^{-1} \sum_z \left(\frac{\partial A_{zt}}{\partial R_{zt}} (R_{zt'} - R_{zt}) \frac{P_{zt}}{P_{zt'} - P_{zt}} \right), \quad (38)$$

where A_{zt} represents the area of cropland in county z during year t , and P_{zt} is a weighted average of crop prices.²⁵

While there is a natural definition of acreage-price elasticities in static models, dynamic

²⁴ I have omitted difference operators to simplify the notation here, but the $\alpha_{0z}(k)$ in equation (35) should still be regarded as a function of differences in parameters of the payoff function (the $\alpha_{0z}(j, k)$ terms).

²⁵ County-level prices are averaged across crops using the same weights as county-level returns – see equation (3.5).

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:

$$\left(\sum_z \sum_\zeta A_{z\zeta}^*(R_{zt}) \right)^{-1} \left(\sum_z \sum_\zeta \left(A_{z\zeta}^*(R_{zt'}) - A_{z\zeta}^*(R_{zt}) \right) \frac{P_{zt}}{P_{zt'} - P_{zt}} \right) \quad (39)$$

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 (38-39), 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.²⁶ Caloric yields are weighted averages of crop-specific yields, using the same weights used for returns (see equation (21)).

5 Results

Table 2 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 .03 – dynamic models generate long-run elasticities on the order of .3.

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. Standard errors are computed allowing for spatial and temporal autocorrelation. See Section A.1 for details.

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 (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 (before differencing out fixed effects). In Table 2, the samples are harmonized across specifications so that differences in results do

²⁶ Recent evidence suggests that intensive crop supply responses are relatively small, at least in the US (Berry and Schlenker, 2011).

Table 2: Long-Run Price Elasticities for Different Specifications

	No Unobs. Heterogeneity		Two Types Per County	
	Acreage	Calorie	Acreage	Calorie
Static model ($\bar{k} = 0$)	0.0156 (0.0050)	0.0152 (0.0048)	0.0256 (0.0079)	0.0263 (0.0079)
Myopic models ($\beta = 0$)				
$\bar{k} = 1$	0.0675 (0.0212)	0.0652 (0.0205)	0.0081 (0.0047)	0.0084 (0.0045)
$\bar{k} = 2$	0.1328 (0.0358)	0.1275 (0.0344)	-0.0016 (0.0218)	-0.0018 (0.0215)
Dynamic models ($\beta = .9$)				
$\bar{k} = 1$	0.6521 (0.3762)	0.6160 (0.3685)	0.2416 (0.0190)	0.2278 (0.0170)
$\bar{k} = 2$	0.3512 (0.3374)	0.3435 (0.3368)	0.3009 (0.1149)	0.2951 (0.1102)

All models feature first differences with instruments. Standard errors in parentheses allow for arbitrary correlation within each year.

not reflect differences in the effective samples.

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,

$$\ln \left(\frac{p_{zt}(k)}{1 - p_{zt}(k)} \right), \quad (40)$$

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

$$\ln \left(\frac{p_{zt}}{1 - p_{zt}} \right) = \ln \left(\frac{\sum_k \mu_{zt}(k) p_{zt}(k)}{1 - \sum_k \mu_{zt}(k) p_{zt}(k)} \right), \quad (41)$$

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 in levels might understate actual long-run responses.

As noted in Section 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.

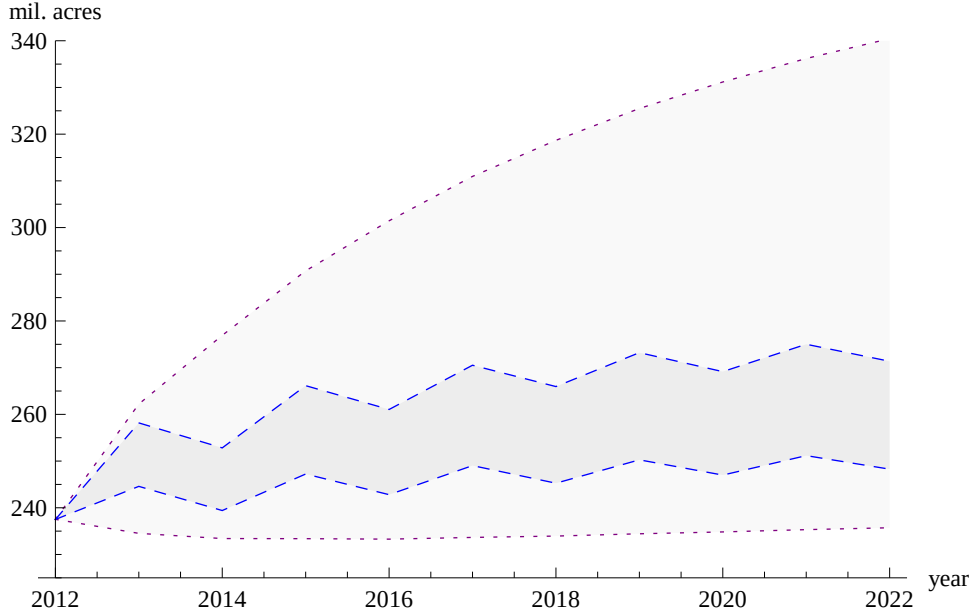
Table 3 provides a more detailed summary of estimation results for dynamic models. For models with unobservable heterogeneity, it is the low field types (those with a lower probability of being planted in crops) which have a higher value of the sensitivity parameter α_R . High field types tend to be in cropland most of the time, and the results reflect that their choice probabilities respond little to variation in expected returns. While the point estimates of long-run elasticities are not significantly different for dynamic models with and without unobservable heterogeneity, the standard errors are considerably smaller for models with unobservable heterogeneity.

Table 3: Dynamic Model Parameter Estimates

State dependence ($\bar{k}$) Field types per county Type	Model Specification					
	1	1	2	2		
	1	2	1	2		
		High	Low		High	Low
α_R	0.0790 (0.0357)	0.0090 (0.0124)	0.3667 (0.0237)	0.0431 (0.0387)	0.0363 (0.0174)	0.2194 (0.0820)
Average intercept (α_{0k})						
$k = 0$	-0.1522 (0.0102)	1.2312 (0.0012)	-2.5792 (0.0045)	-0.0168 (0.0120)	0.7729 (0.0024)	-2.0226 (0.0540)
$k = 1$	-4.7859 (0.0102)	-0.0835 (0.0012)	-5.1790 (0.0045)	-2.3696 (0.0120)	-1.4929 (0.0024)	-1.8249 (0.0540)
$k = 2$	— —	— —	— —	-5.7543 (0.0120)	-1.9172 (0.0024)	-5.0213 (0.0540)
Type's share of fields	1.0000	0.2708	0.7292	1.0000	0.2547	0.7453
Share in crops	0.2503	0.8637	0.0225	0.2523	0.8878	0.0351
Long run acreage- price elasticity	0.6521 (0.3762)	0.2416 (0.0190)		0.3512 (0.3374)	0.3009 (0.1149)	
Long run calorie- price elasticity	0.6160 (0.3685)	0.2278 (0.0170)		0.3435 (0.3368)	0.2951 (0.1102)	
Counties	1975	1975		1975	1975	
County-years	6296	6296		6296	6296	

Intercepts are unweighted averages of the estimated intercepts for all counties. First differences with instruments were used, and $\beta = .9$. Standard errors in parentheses allow for arbitrary correlation within each year. R measured in \$100/acre.

Figure 2: Stability of predictions with and without unobservable heterogeneity



Forward simulated acreage with returns fixed at 2006 levels indefinitely. Dashed lines are 95% confidence interval for total acreage levels in model with unobservable heterogeneity. Dotted line are 95% confidence interval for model without unobservable heterogeneity. Initial distribution of field states in 2012 set equal to the average distribution of field states by county and unobservable type in the sample. Both models feature $\bar{k} = 2$, $\beta = .9$, and were estimated using first differences with instruments.

Putting aside elasticity estimates, there is another important sense in which the results with unobservable heterogeneity are more stable. Figure 2 illustrates the stability of predicted acreage levels for a model with and a model without unobservable heterogeneity. If returns are held constant at 2006 levels indefinitely, the model with unobservable heterogeneity predicts relatively stable acreage levels going forward. In contrast, the model without unobservable heterogeneity allows for relatively large increases in acreage within its 95% confidence intervals.

It makes sense that models with unobservable heterogeneity would make more stable land use predictions. For a given observable type of field, if we observe considerable amounts of both cropland and non-cropland and a non-trivial amount of switching between the two, the only way to rationalize the data without unobservable heterogeneity is for landowners to be somewhat close to indifferent about whether or not to switch land uses. Then, if returns shift dramatically, the model will predict a massive shift in land use behavior. In contrast, the model with unobservable heterogeneity allows for some landowners to be very happy to stay in cropland and others to be very happy to stay out of cropland, and both types may be relatively unresponsive to changes in returns since they are far from indifferent on average.

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.

The US biofuels mandate is met mostly by corn ethanol, and the corn used to produce this ethanol corresponds to a staggering 5% of worldwide caloric production of major crops.²⁷ 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.

Following RS, 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), my static supply elasticity of 0.03 implies a 44% increase in prices and a 1.1% increase in crop acreage.²⁸ In contrast, my dynamic long-run elasticity of 0.3 implies a 9.7% increase in prices and a 2.9% increase in crop acreage. In other words, taking dynamics into account when estimating supply leads to a 160% larger land use effect and a 78% smaller price increase in the long run.

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. Moreover, it should be noted that the relatively small price increase following from the dynamic elasticity estimate is based on plugging a long-run

²⁷It 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.

²⁸I 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.

supply elasticity into a static equilibrium calculation. Even if the long-run price effect of the US biofuels mandate is as small as this 9.7% figure suggests, it is possible that the mandate had a much larger short run price effect, consistent with the spike in prices in 2008.

RS estimate a global calorie-price elasticity around 0.1, but they estimate an acreage-price elasticity for the US which is remarkably similar to my estimate of 0.3. This is surprising given that their estimate is based on a static model while my static models deliver much smaller elasticities. However, RS’s estimate is based largely on much older data, and they discuss the possibility that their supply elasticity estimates are picking up an endogenous policy response – i.e., policies incentivizing farmers to set land aside may arise when output prices are low. This endogenous policy response has arguably weakened over time, making the effective static elasticity lower now. Thus, it may be the case that their supply elasticity, which is based on much older data, is picking up an older static relationship which happens to coincide with the new long-run elasticity.

Another possible source of the difference between RS’s and my static estimates is the difference in instrumental variables used. Unfortunately, RS’s instruments based on lagged yield shocks (or weather variation) are too weak to deliver reliable estimates with my short panel. Notice that the explanatory variable model in my model is proportional to $R_{crops,t+1} - R_{crops,t}$. While this difference in expected returns will certainly be strongly correlated with weather outcomes during year t , these weather outcomes are also correlated with unexpected changes in the value function ε_t^V , making time- t weather and yield shocks invalid instruments. Weather outcomes from earlier years are plausibly valid instruments, but I find that they are too weak to generate estimates that can be taken seriously.

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 when 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.

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.
- Bajari, Patrick, C. Lanier Benkard, and Jonathan Levin**, “Estimating Dynamic Models of Imperfect Competition,” *Econometrica*, 2007, *75* (5), 1331–1370.
- 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.
- 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.
- 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.
- , **Robert A. Miller, Seth Sanders, and Jeffrey Smith**, “A Simulation Estimator for Dynamic Models of Discrete Choice,” *Review of Economic Studies*, April 1994, *61* (2), 265–89.
- 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.
- 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.
- 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.
- Nerlove, Marc**, “Estimates of the Elasticities of Supply of Selected Agricultural Commodities,” *Journal of Farm Economics*, 1956, 38 (2), 496–509.
- 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,” *Quantitative Economics*, 2013, 4 (1), 149–155.
- 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.
- Roberts, Michael J. and Wolfram Schlenker**, “Identifying Supply and Demand Elasticities of Agricultural Commodities: Implications for the US Ethanol Mandate,” *American Economic Review*, 2013, 103 (6), 2265–95.
- 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.

A Technical appendix (for online publication)

A.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. Conley’s treatment of 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.

The asymptotic theory developed in these papers relies on covariance matrix estimators which use weights which decline as the distance (and/or time duration) between observations increases. Furthermore, the weights change as the sample size increases. In practice, it is 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 ζ -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 my main specification (first-differences with instruments), fitted moments can be computed as follows:

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

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,

$$K(s_1, s_2) = \begin{cases} 1 & \text{if } t_1 = t_2 \\ .5 & \text{if } |t_1 - t_2| = 1 \\ 0 & \text{otherwise} \end{cases}$$

This kernel allows for arbitrary correlation within the cross-section, which is important because farmers everywhere receive similar price shocks and therefore are likely to have correlated expectational error terms ε^V . Because the dependent variables in my regression use choice probabilities from the time t and $t + 1$, it also makes sense to allow for some correlation over time. I use the Bartlett kernel in the time dimension (i.e., weights dropping off linearly) to ensure that $\hat{V}$ is positive semi-definite.

Table 6 includes standard errors for estimates computed with and without corrections for autocorrelation. In other tables, I always present standard errors accounting for autocorrelation.

A.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:

$$\mu_{LRE} = \frac{1}{N_{sim}} \sum_{l=1}^{N_{sim}} LRE(\alpha_l)$$

$$SE_{LRE} = \sqrt{\frac{1}{N_{sim} - 1} \sum_{l=1}^{N_{sim}} (LRE(\alpha_l) - \mu_{LRE})^2}$$

where α_l represent pseudorandom draws from the estimated asymptotic distribution of α , and $N_{sim} = 1000$ is the number of simulations.

A.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 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 14.

A.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 roughly inversely proportional to the estimated standard error of the constructed dependent variable.

The variance of the log of an estimated probability ($\ln(\hat{p})$) is largest for small probabilities, so the sampling variance of the dependent variable $Y_{z,\zeta,k,t}$ will tend to be dominated by the smallest choice probability used to construct it. Let $\underline{p}_{z,\zeta,k,t}$ be the smallest choice probability among those used to construct $Y_{z,\zeta,k,t}$ (see equation (16)). I construct weights for the regression equation as follows:

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

where $\underline{p}_{z,\zeta,k}$ is the mean value of $\underline{p}_{z,\zeta,k,t}$ across t , and $\hat{N}_{z,\zeta,k}$ is the average number of fields in field state k in county z of type ζ . I use the same weight $w_{z,\zeta,k}^Y$ for all periods t .

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

A.5 First stage choice probability estimation

For specifications with unobservable heterogeneity, choice probabilities are estimated by alternating between the expectation step (equation (26)) and maximization step (equations (27-29)) until the change in the likelihood function between full iterations is less than 10^{-8} for ten successive iterations. This first-stage estimation is run separately for each specification and US state, allowing the first-stage estimation to benefit substantially from parallelization.

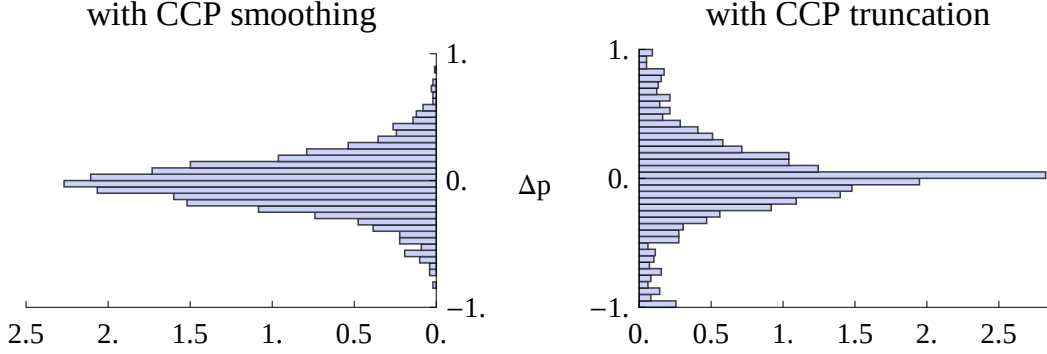
As described in Section 4.1, I smoothed choice probability estimates across counties within each state using weights proportional to the inverse square of the distance between counties. My smoothing weights are small, but only a tiny amount of smoothing is needed to avoid CCP estimates of ones and zeros, which present a technical problem for the Hotz-Miller inversion.

Another way to avoid degenerate CCP estimates is to truncate them, or replace any frequencies of zero with an estimate which is very close but slightly larger than zero.²⁹

I experimented with CCP truncation, but found that it led to considerably less stable CCP estimates than CCP smoothing. While CCP truncation allows for CCP's to be estimated within each county in principle, there is often insufficient data within each county to have

²⁹ An earlier version of the paper included estimates from both smoothing and truncation.

Figure 3: Stability of CCP estimates



PDF histograms of time-differenced choice probabilities, $p_{z,\zeta,t+1}(crops,k) - p_{z,\zeta,t}(crops,k)$, over all counties for $t = 2010$, $k = 2$, and $\zeta = 2$. Other field states and time periods display a similar pattern in that truncated CCPs are considerably less stable.

reasonable CCP estimates for both unobservable types and all field types. When there are effectively no observations for a given (ζ, k) -pair for a given county z and year t (or, more formally, the EM algorithm predicts that all fields in field state k have very low probability of being in type ζ), it makes more sense to base the estimate of $p_{z,\zeta,t}(k)$ on neighboring counties which actually have fields of type ζ in state k .

Just as Pakes et al. (2007) argue, smoothing can reduce the sampling variance of CCP estimates. Indeed, I find that year-to-year changes in choice probabilities for a given type and field state (z, ζ, k) are much more stable with smoothing than with CCP truncation. As Figure 3 illustrates, truncated choice probabilities sometimes go from 0 to 1 or vice versa, but choice probabilities with smoothing do not display such extreme changes.

Another apparent virtue of the CCP smoothing over CCP truncation is that it dramatically improves the convergence properties of the EM algorithm. This is ironic, as smoothing takes us away from the maximum likelihood framework and removes the theoretical guarantee that the EM algorithm will converge. However, with smoothing, I find that the EM algorithm not only always converges but also always converges to the same CCP estimates. With truncation, the EM algorithm converges to many different local maxima of the likelihood function depending on the starting values.

With CCP smoothing, I run the EM algorithm with five randomly selected starting values for each state. For each state, all five trials always converge to the same CCP estimates. With CCP truncation, I run the EM algorithm forty separate times for each county. On average, 14 out of the 40 trials converge to the highest value of the likelihood function attained for the county. However, for over 24% of counties, the highest value of the likelihood function attained was attained by only one trial, which casts doubt on whether 40 trials is enough to reliably find the CCP estimates which globally maximize the likelihood function.

I do not report results based on CCP truncation since these CCP estimates cannot be considered reliable. However, elasticities estimated based on my truncated CCP's are not significantly different than my results with CCP smoothing.

A.6 Generalizing the regression equation derivation

In this section, I discuss the generality of the derivation of the regression equation (15). In particular, I consider three aspects of the model presented in Section 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 pair implying a different version of equation (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 -period sequences of CCPs. To extend my regression equation construction to s -period dependence, equation (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 regression equation derivation 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 (8)), and in equation (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 ($v(j, k)$) and conditional choice probabilities. As shown below, the existence of an additively separable equation analogous to equation (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 $v_t(j, k)$:

$$\forall j \in \mathbf{J} : p_t(j, k) = \Pr(\forall j' \in \mathbf{J} : v_t(j, k) + \nu_{jt} \geq v_t(j', k) + \nu_{j't}) . \quad (42)$$

Normalizing $v_t(j, k)$ to zero for some j , equation (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.³⁰

One value of $v_t(j, k)$ 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 $\phi_j^{-1}(\mathbf{p}_t(k)) = v_t(j, k) - v_t(J, k)$, where $\mathbf{p}_t(k) = \{p_t(j, k)\}_{j \in \mathbf{J}}$.

Result 2 shows that that equation (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), the following proof is simpler than theirs.³¹

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

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

Proof. Define

$$S(v) \equiv \int \max\{v_1 + \nu_1, v_2 + \nu_2, \dots, v_J + \nu_J\} dF(\nu) . \quad (43)$$

For any real number c ,

$$\max\{v_1 + \nu_1 - c, v_2 + \nu_2 - c, \dots, v_J + \nu_J - c\} = \max\{v_1 + \nu_1, v_2 + \nu_2, \dots, v_J + \nu_J\} - c .$$

It follows that

$$S(v) = S(v_1 - v_J, v_2 - v_J, \dots, 0) + v_J .$$

³⁰Being 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, 2013).

³¹I thank Steve Berry for suggesting this proof strategy.

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

$$\bar{V}_t(k) = S(v_t(1, k) - v_t(J, k), v_1(2, k) - v_t(J, k), \dots, 0) + v_t(J, k). \quad (44)$$

Using Proposition 1, substitute $\phi_j^{-1}(p_t(k)) = v_t(j, k) - v_t(J, k)$ into equation (44):

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

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 (11) is a particularly simple special case of Result 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 (11) is preserved, maintaining the possibility of constructing a linear regression equation.

B Data appendix (for online publication)

B.1 Land cover panel

The accuracy of the Cropland Data Layer 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.³² 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 are classified as developed land, water, or protected land (in any year) are excluded from the data set. This means that non-cropland includes mostly grassland, forests, and shrub land.

Points are 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

³²CDL accuracy data is available at <http://www.nass.usda.gov/research/Cropland/sarsfaqs2.html>

neighbor interpolation. All coordinate conversions and spatial merges were done with ArcGIS 10.2.

B.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 within each county to form the county-level weather variables which are used in the yield regressions.

B.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

Table 4: 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

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

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

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.

B.4 Expected prices

All reference prices are based on contracts for December delivery, except for soybean prices, which are based on November delivery contracts.

To deal with differences in planting seasons, I actually estimate two versions of equation (19), 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., $P_{czt}^m = P_{cst}^{m,sp}$).³³

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

B.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 crop c in county z if (c, z) is included in the yield regression (as described in Section B.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,

$$\frac{\sum_{c \in \mathbf{C}_z} A_{crt}}{\sum_{c \in \mathbf{C}} A_{crt}} > 9.$$

where r is the state county z belongs to.

Furthermore, thirteen states were excluded from the analysis 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.

³³The 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.

Figure 4: Map of sample counties

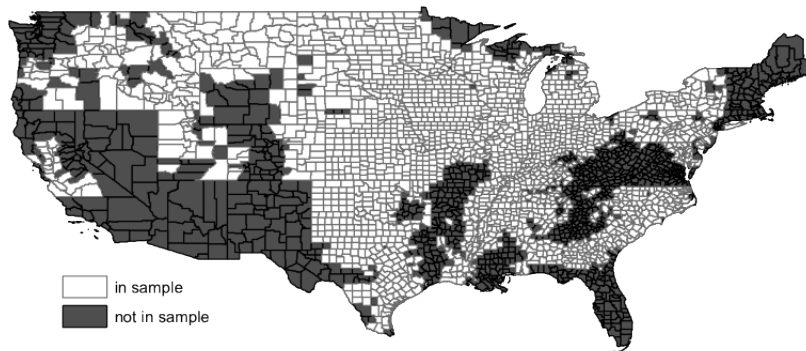


Table 5: Long-run Elasticities for Different Measures of Returns

Acreage elasticity	0.3009 (0.1149)	0.3792 (0.1330)	0.1315 (0.1556)	0.1649 (0.1652)
Caloric elasticity	0.2951 (0.1102)	0.3773 (0.1303)	0.1322 (0.1568)	0.1676 (0.1686)
Price forecasts based on	planting season futures		futures from prev. fall	
Costs per acre are	proportional to yields	flat within region	proportional to yields	flat within region

All models feature two unobservable types, two periods of state dependence, first differences with instruments, and $\beta = .9$. Standard errors in parentheses allow for arbitrary correlation within each year.

According to the 2007 Census of Agriculture, the counties remaining in my sample account for over 90% of US cropland. Figure 4 presents a map of these counties.

C Sensitivity analysis (for online publication)

C.1 The expected returns measure

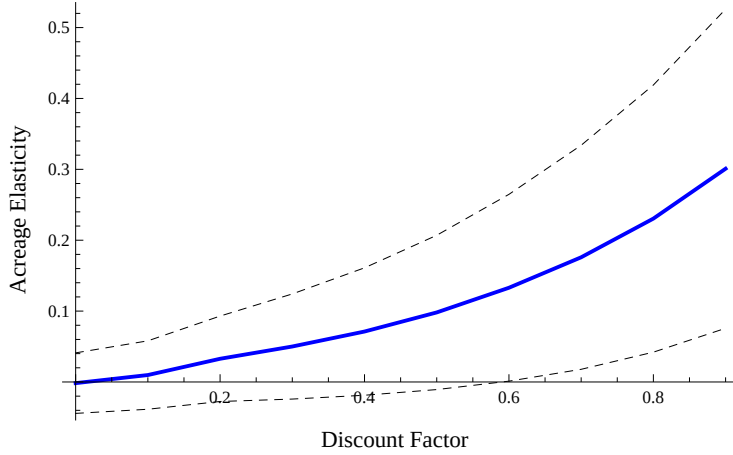
Constructing the measure of expected returns involves two substantive assumptions. First, expected prices were forecasted using futures prices around the beginning of planting season as explained in Section B.4. However, farmers may effectively commit to planting crops substantially before this time. 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.

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 some of the 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.

Table 5 shows how these two assumptions behind the construction of returns impact long-

Figure 5: 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.

run elasticity estimates. Assumptions on the timing of the decision make a more substantial difference although the difference is not statistically significant.

C.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 estimate the model for different discount factor imputations, and as Figure 5 illustrates, I find that the relationship between long-run elasticity estimates and the discount factor tends to be increasing and convex.

C.3 Different estimation approaches

Table 6 illustrates how regression weights and differencing affect the estimates, and presents standard errors with and without allowing for autocorrelation.

While the weights and differencing strategy seem to make relatively little difference, the impact of autocorrelation is large, reflecting considerable spatial autocorrelation across counties.

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 4.3 may be a relatively small issue in practice.

Table 6: Weights, Differencing, and Correlation

Weights	Std. Errors	Regression Approach		
		FE	FD	FDIV
yes		0.2537	0.2234	0.3009
	no correlation	(0.0124)	(0.0108)	(0.0159)
	spatial HAC	(0.1859)	(0.1232)	(0.1149)
no		0.2554	0.2435	0.3434
	no correlation	(0.0102)	(0.0094)	(0.0125)
	spatial HAC	(0.2404)	(0.1806)	(0.2191)

Regression approaches are fixed effects, first differences, and first differences with instruments. All models feature two unobservable types, two periods of state dependence, and $\beta = .9$.

Table 7

LAND COVER CLASSIFICATIONS					
CDL Classification	My Classification	land cover %	Developed, High Intensity	CDL Classification	My Classification
Grassland/Herbaceous	other	18.2621	Developed, High Intensity	excluded	0.200302
Deciduous Forest	other	11.8835	Barley	crops	0.142016
Shrub/Scrub	other	11.4498	Sunflower	crops	0.10755
Corn	crops	8.22904	Dry Beans	crops	0.0972723
Evergreen Forest	other	8.00622	Sugarbeets	crops	0.094715
Soybeans	crops	6.4165	Durum Wheat	crops	0.0943772
Pasture/Hay	other	5.30431	Oats	crops	0.0936696
Developed, Open Space	excluded	3.84439	Canola	crops	0.0902277
Woody Wetlands	other	3.58478	Peanuts	crops	0.0779722
Winter Wheat	crops	3.10737	Potatoes	crops	0.0705899
Pasture/Grass	other	2.85289	Sod/Grass Seed	other	0.0630629
Fallow/Idle Cropland	other	2.07242	Almonds	crops	0.0461432
Open Water	excluded	1.76229	Peas	crops	0.0430874
Non-alfalfa Hay	other	1.68951	Millet	crops	0.0408839
Developed, Low Intensity	excluded	1.51962	Grapes	crops	0.0404336
Alfalfa	other	1.34933	Rye	crops	0.0388735
Cotton	crops	1.32377	Lentils	crops	0.0341772
Herbaceous Wetlands	other	1.29701	Apples	crops	0.028162
Spring Wheat	crops	1.20907	Walnuts	crops	0.0275508
Mixed Forest	other	0.739771	Dbl. Crop WinWht/Sorghum	crops	0.0236586
Barren Land	other	0.690379	Pecans	crops	0.0235943
Developed, Medium Intensity	excluded	0.5409	Dbl. Crop WinWht/Cotton	crops	0.0235139
Sorghum	crops	0.440813	Sweet Corn	crops	0.0219699
Dbl. Crop WinWht/Soy	crops	0.418087	Aquaculture	excluded	0.0189462
Rice	crops	0.243406	Clover/Wildflowers	other	0.0159869
Percentages are for counties in my sample in 2011. Only land cover classifications with at least 950 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

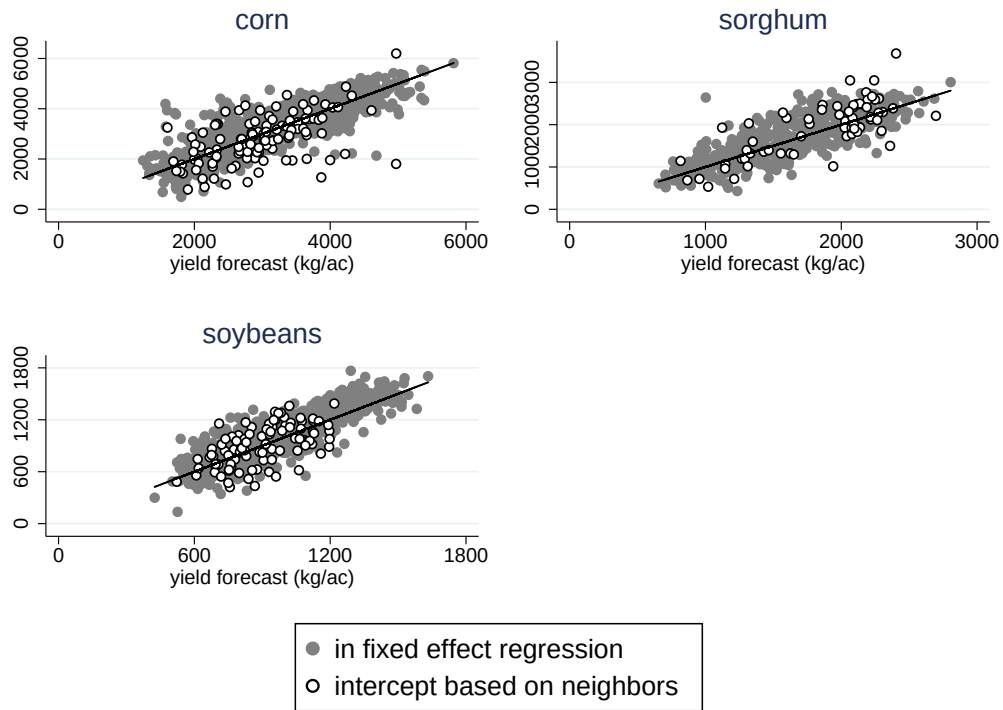
VARIABLES	(1) region 1	(2) region 4	(3) region 5	(4) region 6	(5) region 7	(6) region 8	(7) region 9
DD10	0.00014 (0.00021)	-0.00088*** (0.00010)	0.00047*** (0.00014)	-0.00017*** (0.000058)	0.00068*** (0.000082)	0.00076** (0.00029)	0.00040*** (0.000058)
DD30	-0.0091*** (0.0011)	-0.00050 (0.00041)	-0.013*** (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)	0.00099 (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

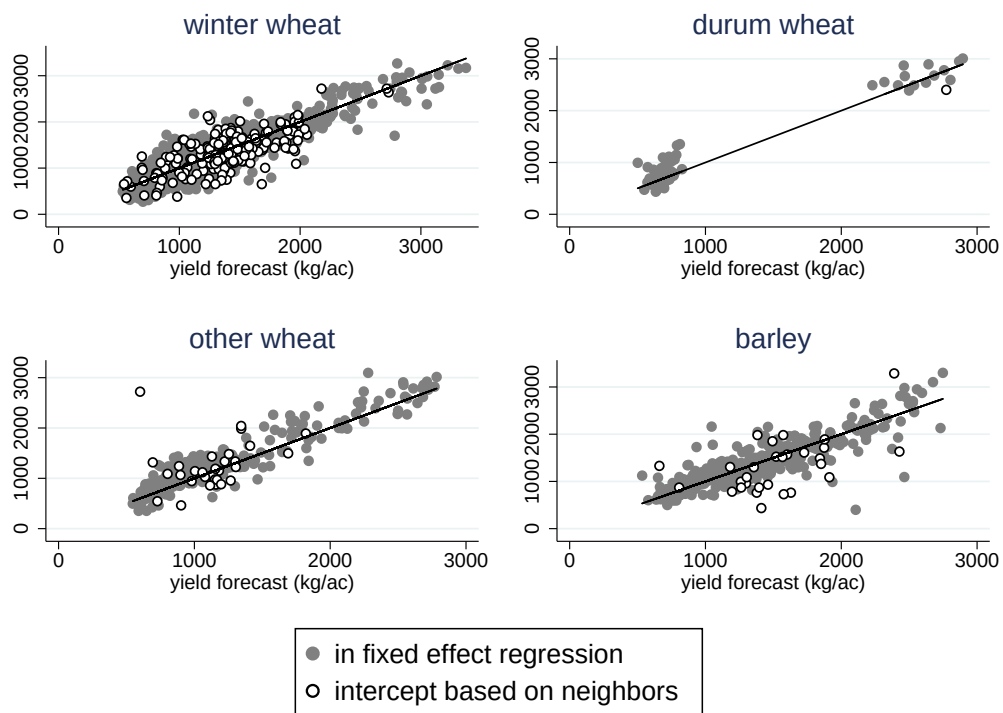
Dependent variable: log yield. Specifications include county fixed effects and linear time trend

Figure 6: Validation of expected yields, selected crops



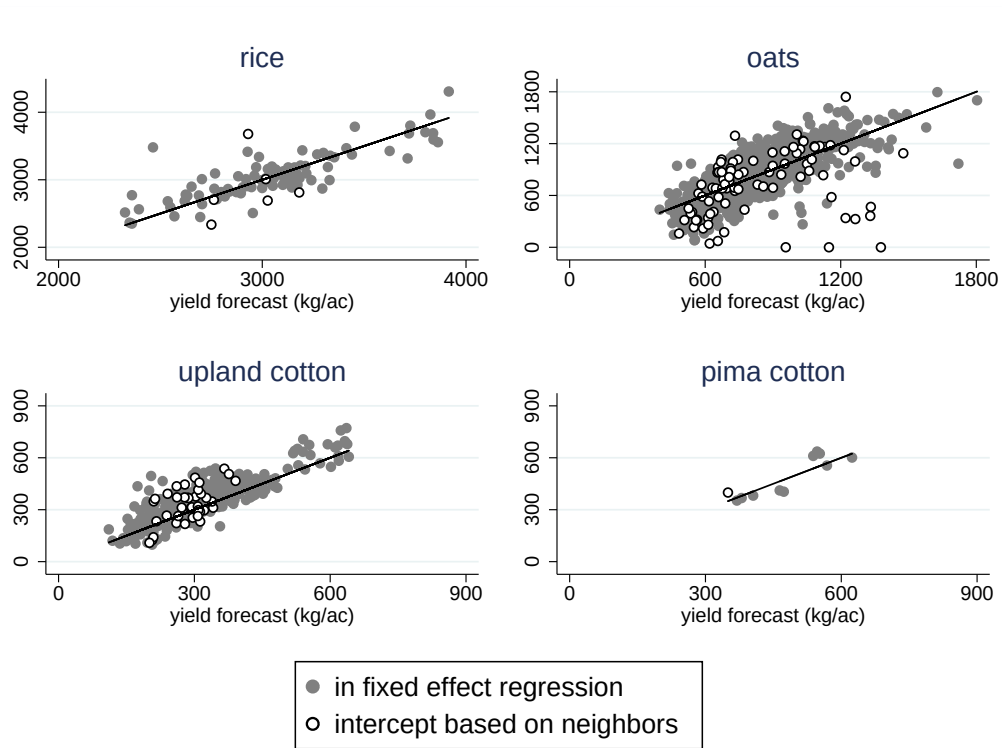
Each point corresponds to a county, with the x-axis indicating expected yield forecasts for that county, averaged over 2006-2011; the y-axis represents the average actual yield for that county, averaged over the same period (reported by the National Agricultural Statistics Service).

Figure 7: 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 8: 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).