Let Them Eat Switchgrass? Modeling the Displacement of Existing Food Crops by New Bioenergy Feedstocks*

Soren Anderson†
Michigan State University and NBER

Chenguang Wang‡
Michigan State University

Jinhua Zhao§
Michigan State University

November 20, 2012

PRELIMINARY: PLEASE DO NOT CITE WITHOUT PERMISSION

*For helpful comments and suggestions, we thank Bruno Basso, Roy Black, Ryan Kellogg, Philip Robertson, Scott Swinton, and seminar participants at Michigan State University, Iowa State University (Bioenergy Camp), University of Illinois (Heartland Environmental and Resource Economics Conference), and the University of California Energy Institute. We thank Ryan Kellogg, Michael Roberts, and Wolfram Schlenker for sharing climate data. We gratefully acknowledge financial support from Michigan State University's Rackham Research Endowment. Any remaining errors are our own.

†Corresponding author. Email: sta@msu.edu. Phone: 517-355-0286. Web: www.msu.edu/∼sta.

‡Email: wangch28@msu.edu.

§Email: jzhao@msu.edu. Phone: 517-353-9935. Web: www.msu.edu/∼jzhao

Abstract

We propose an econometric model of crop choice in which a farmer's profit-maximizing crop depends on physiological crop attributes interacted with local soil and climate characteristics. We estimate the model using county-level data on average crop choices for 1986–2008. We find that attributes of chosen crops correlate as expected with geographic variation in soil and climate characteristics. For example, crops that use water efficiently tend to be grown in dry regions, while crops with long growing seasons tend to be grown in warmer climates. The model allows us to forecast, based on revealed historical crop choices, the adoption of two leading bioenergy crops, switchgrass and miscanthus, that do not yet exist in the historical data but whose attributes are known. At the scale of production mandated by the federal Renewable Fuel Standard, we find that both switchgrass and miscanthus would displace a wide range of major food crops. Driven by their relative tolerance for cold, these new crops would tend to concentrate in the northern plains, with disproportionately large impacts on alfalfa, barley, sunflowers, and wheat and somewhat milder impacts on corn, cotton, sorghum, and soybeans. These results could be used to validate the findings of simulation models or to help parameterize a model of U.S. crop supply, while our overall approach could easily be extended to include additional crops and countries.

JEL classification numbers: Q1, Q2, Q4

Key words: crop choice, agricultural supply, bioenergy feedstocks

1 Introduction

All crops need sun, water, and nutrients to grow, but different crops require these inputs to varying degrees. Thus, a crop that grows well in one location may be more costly to grow somewhere else, as sunshine, rainfall, and soil quality all vary across locations. Crop choices will therefore tend to correlate geographically, as profit-maximizing farmers match crop attributes to soil and climate conditions. For example, farmers working in arid regions will tend to choose crops that require little water, while those working in regions with low soil fertility will tend to choose crops with minimal nutrient needs, all else equal.

This matching of crop physiology to land and climate characteristics is important for predicting the supply costs and side-effects associated with growing bioenergy feedstocks, such as switchgrass, sugar cane, and miscanthus. The production of these feedstocks is subsidized both directly through tax credits and indirectly through renewable fuel standards in the United States, the European Union, Brazil, and other countries. The cost of producing these feedstocks will depend both on the global availability of land suitable for their production, as well as on the demand for existing food crops that these bioenergy feedstocks would potentially displace. Demand for food is insensitive to price changes (Roberts and Schlenker, 2010a). Thus, if new bioenergy feedstocks compete strongly for land with existing food crops, the cost of producing these feedstocks will be high, and government policies or rising energy prices that boost demand for feedstocks will tend to drive up food prices.

To address these issues, we propose a discrete-choice econometric model of crop supply in which a farmer's profit-maximizing crop choice depends on physiological crop attributes, such as drought resistance and nutrient demand, interacted with local soil and climate characteristics, such as annual rainfall and soil fertility. Because we model crop choices in terms of generic crop attributes, our approach allows us to forecast statistically the supply of new crops that do not yet exist in the historical data but whose physiological attributes are known. We estimate this model using county-level data on average crop choices for 1986–2008. As expected, we find that observed physiological attributes explain much of the geographic variation in crop choices. For example, crops that use water efficiently tend to be grown in dry regions, while crops with long growing seasons tend to be grown in warmer climates.

We use our model to forecast the adoption of two leading bioenergy feedstocks, switchgrass and miscanthus, based on their physiological attributes. At the scale of production mandated by the U.S. Renewable Fuel Standard (RFS), we find that both feedstocks would displace a wide range of food crops. Thus, unless production of these new bioenergy crops can be confined to land that is not currently being used for cultivation, policies that mandate the production of cellulosic ethanol will inevitably displace existing food crops, leading to higher food prices. Our model predicts that both feedstocks would tend to concentrate in the northern plains, largely driven by their relative cold tolerance, with disproportionately large impacts on alfalfa, barley, sunflowers, and wheat and somewhat milder impacts on corn, cotton, sorghum, and soybeans. Miscanthus would likely have milder impacts on world food prices overall, however, given its higher projected yields. In future work, we plan to extend our model to incorporate other land uses, including rangeland, forests, and pasture, to explore possible interactions between climate change and bioenergy crop production, and to explore how adopting new crops or modifying the attributes of existing crops might contribute to climate change adaptation in the agricultural sector.

This paper is related to an immense economic literature, with roots stretching back many decades, modeling various aspects of agricultural supply. One recent strand of this literature builds bottom-up structural models of crop choice and output. These models typically specify crop yields and production budgets that vary exogenously across locations, using linear programming techniques to calculate the profit-maximizing crop for each location. Less frequently, crop choices are based on physiological plant growth functions, developed by crop and soil scientists, that directly tie yields for specific crops to soil and climate characteristics, allowing modelers to predict how crop choices might respond to climate change. Such models can also predict the supply of new crops, such as switchgrass and miscanthus, based on what is already known about the growth of these plants. In all of these models, the economic behavioral components are typically assumed or calibrated, rather than estimated, and there is typically no accounting for the unobserved costs and benefits of adopting a particular crop, whether systematic (i.e., crop fixed effects) or idiosyncratic (i.e., prediction errors). Thus, when the models make implausible predictions, modelers often impose ad-hoc constraints on crop choices in particular regions, so that predicted choices match more closely what is observed in practice—obviously, an undesirable feature of a model whose purpose is to predict crop choice.[1]

A second strand of literature estimates econometric models of crop choice and agricultural supply. These models range from parsimonious crop choice models based on reduced-form profit functions (Seo and Mendelsohn, 2008) to more complex models that embed explicit yield and input demand functions (Timmins, 2002). Like bottom-up simulation models, these econometric models have the ability to predict crop choices in specific locations, which allows them to forecast, for example, how cropping patterns might respond to climate change. The main advantage of these models is that their parameters are estimated directly from revealed choices. One disadvantage, however, is that these models identify choice parameters that are specific to particular crops, making it difficult or impossible to predict behavior for new crops that do not already exist in the historical data, or how cropping patterns might respond to changes in crop attributes due to plant breeding programs.[2]

Our model can be viewed as a hybrid of these two approaches. Like bottom-up simulation models that base crop choices on detailed yield functions, we incorporate information on plant physiology that has accumulated over decades in the crop and soil science literature and treat

[1]Notable examples of models that impose constraints on crop choices and land use include POLYSYS, as well as FASOM, which the EPA has used to calculate the impacts of the federal RFS on life-cycle carbon emissions. Egbendewe-Mondzozo, Swinton, Izaurralde, Manowitz, and Zhang (2011) and Khanna, Chen, Huang, and Onal (2011) develop models that partially relax the assumption of inelastic supply at the constraints, respectively, by incorporating information on empirical supply elasticities and by calibrating downward-sloping yield functions.

[2]Another potential disadvantage, also present in our analysis, is that these models, which contain relatively few parameters, represent stylized models of crop production, potentially overlooking important behaviors, such as crop rotation.

this information as data. We focus only on the most salient physiological attributes, which are either intrinsic to the plant (such as whether the plant uses C3 or C4 photosynthesis), have been quantified numerically using statistical methods (such as the efficiency with which the plant uses water), or are understood well enough to be ranked qualitatively (such as the plant's salt tolerance).[3] Like other econometric choice models, however, we estimate the behavioral parameters of our model by matching observed and predicted crop choices. Importantly, we do not estimate crop-specific choice parameters. Instead, we define crops according to their observed physiological attributes, and we estimate choice parameters for these crop attributes interacted with soil and climate characteristics. Thus, unlike existing econometric models, we are able to forecast the supply of new crops that do not already exist in the historical data but whose physiological attributes are known.[4]

Our paper is also related to a growing economics literature that uses hedonic methods to estimate the agricultural land values as a function of soil and climate characteristics, often to value the potential damages associated with climate change.[5] The seminal paper in this literature is Mendelsohn, Nordhaus, and Shaw (1994), while a string of more recent papers have continued to refine and extend these methods (Schlenker, Hanemann, and Fisher, 2005, 2006; Timmins, 2006; Massetti and Mendelsohn, 2011). Of course, observed land values depend on equilibrium crop prices, and large changes in climate could seriously disrupt this equilibrium. Moreover, hedonic models alone cannot predict the supply of any

[3]While there is an extensive scientific literature relating crop yields to various inputs, on which we draw to quantify physiological crop attributes, these studies are scattered across a wide range of locations and time periods, and use a wide range of methods. Thus, we believe there is considerable room for further research that quantifies crop attributes rigorously and consistently using modern quasi-experimental empirical methods, as in Schlenker and Roberts (2009) and Roberts and Schlenker (2010b), who estimate the impact of random temperature fluctuations on corn, soybean, and cotton yields.

[4]Our approach parallels that of Timmins (2002) and Asrat, Yesuf, Carlsson, and Wale (2010), in which crop choices are allowed to depend systematically on observed attributes of particular crop varieties; the first study is based on actual crop choices in Brazil, while the second is based on stated preferences for a sample Ethiopian households. Our approach differs in that we measure multiple physiological attributes of the crops themselves to predict crop choices, whereas these other papers focus on crop varieties and incorporate information on just one or two varietal attributes to predict choice of crop variety.

[5]A complementary literature uses hedonic methods to estimate the price premium for seed varieties with desirable attributes, which can be used to value marginal changes in crop attributes (Dalton, 2004; Ekanem and Sundquist, 1993).

particular crop, let alone a new bioenergy crop. Thus, for large changes in climate, or to model bioenergy feedstock supply, information about the underlying crop supply function is required.

Finally, our paper is related to a new and expanding economic literature that specifically studies the impacts of bioenegy feedstock production and related policies on land, food, and fuel markets using a range of methods to model agricultural supply. For example, Chen, Huang, Khanna, and Önal (2011) estimate the effects of renewable fuel standards and tax credits on cropland allocation, food prices, and energy prices using a dynamic general equilibrium model of food and energy markets, which relies at its core on a bottom-up structural model of crop choice of the type described above.[6] In contrast, Holland, Hughes, Knittel, and Parker (2011) rely on a model of feedstock supply in which crop choices are fixed to estimate the impacts of carbon pricing, renewable fuel standards, and low-carbon fuel standards on fuel prices, quantities, and carbon emissions. Feedstock supply derives from the current cost and availability of existing crops, crop residues, and municipal waste streams, with transportation costs determined endogenously by optimizing the sizes and locations of biorefineries. Finally, Roberts and Schlenker (2010a) estimate the impacts of the corn-based ethanol standard on global food prices using reduced-form linear models of global food supply and demand, which they identify using weather-induced shocks to global crop yields as instruments for prices. Thus, models in the literature reflect a range of assumptions with respect to (1) the scope of behaviors being modeled explicitly, from highly detailed structural models to reduced-form models of aggregate supply, (2) the extent to which the modeling results reflect land-use change, either explicitly or implicitly, (3) whether the models are capable of handling new crops, and (4) the methods used to choose parameters, including calibration, econometric estimation, or a combination of the two.

Our paper makes several contributions to this extensive literature. First, we estimate crop choice as a function of a crop's physiological attributes, which allows us to forecast the supply

[6] They also review other studies that use similar methods, including a cogent discussion of some of weaknesses in the literature's use of these types of models.

of new crops that do not yet exist in the historical data. Previous econometric studies have applied similar methods to analyze the adoption of new crop varieties featuring one or several quantifiable attributes, but never the adoption of entirely new crops, as far as we know. Second, in contrast to virtually all papers on the supply of new bioenergy crops, which use calibrated numerical simulation models, the parameters of our model are estimated directly from observed crop choices. Thus, we argue that our model has the potential to predict some elements of farmer behavior more accurately. While existing simulation methods explicitly model a broader range of behaviors and outcomes, the parameters of our model implicitly reflect many of these same behaviors. For example, if farmers typically add low-cost chemical fertilizers when growing crops in nutrient-depleted soils, then this behavior will be captured in our model as a weak estimated correlation between a soil's nutrient content and the nutrient demands of chosen crops. Third, we control explicitly for unobserved crop and county fixed effects, and any remaining errors are relegated to our model's residual terms, which we present transparently so that readers can evaluate our model's performance. We prefer this econometric approach to methods that impose ad-hoc, location-specific constraints on crop choices to match observed and predicted outcomes, which may only obscure serious errors in these models. Fourth, while our model captures essential information about a crop's production function in the form of the crop's physiological attributes, we do not need to estimate the production function directly to model crop choice. Thus, our model is more transparent, easier to implement with limited data, and computationally simpler than many existing numerical simulation and structural econometric models, and we present a conceptual model that clarifies precisely the nature of our model's assumptions.

Finally, we note that our empirical approach is analogous to a large economics and marketing literature modeling household demand for differentiated products, such as cars and breakfast cereals, as a function of product attributes, such as a car's size or a cereal's fiber content (Berry, Levinsohn, and Pakes, 1995; Nevo, 2000). In these models, a household is assumed to choose the product that yields the highest utility, which depends on product attributes interacted with household characteristics. That is, for example, large families derive greater utility from a car's interior volume, and so large families tend to choose bigger cars. Thus, product choices correlate across households in a way that matches product attributes and household characteristics, allowing modelers to predict demand for new products (Petrin, 2002). While previous studies have applied similar models to the choice of particular crop varieties, we show that the same empirical approach can be used to forecast the supply of entirely new and distinct crops.

The remainder of this paper proceeds as follows. Section 2 presents our conceptual and econometric models. Section 3 discusses important issues in plant growth and crop production, which guide the specification of our empirical model. Section 4 describes the data we use to estimate our model. Section 5 discusses identification and presents our estimation results. Section 6 uses these results to forecast the supply of bioenergy feedstocks and predict which food crops they will likely displace. Section 7 then concludes.

2 Model

To motivate our econometric analysis below, we develop a conceptual model of a profit-maximizing farmer in which, conditional on growing a given crop, expected profits are a linear function of crop attributes interacted with soil and climate characteristics. Given a distributional assumption on the unobserved component of profits, this model then leads naturally to a discrete-choice empirical model of the farmer's profit-maximizing crop choice that can be estimated econometrically using county-level aggregate data. We impose a fair bit of structure on the underlying conceptual model to keep the econometric model simple, to minimize data requirements, and to facilitate our simulations below. Future research could relax these restrictions.

2.1 Crop production

Consider an agricultural landowner that maximizes her land's value by choosing the most profitable crop and growing it optimally. Formally, let the expected per-acre profit from planting crop j on a particular fixed-quantity parcel of land be given by the following expression:

$$\pi_j = p_j \mu_j f(\lambda_j z + x) + \theta_j - rx, \tag{1}$$

where: p_j is the expected end-of-season price for the crop's marketable output in the relevant units; $y_j \equiv \mu_j f(\cdot)$ is per-acre yield in the relevant units, with $f(\cdot)$ a strictly increasing, strictly concave, continuous, and differentiable function of fixed and variable inputs that directly influence yields, and μ_j a crop-specific constant, as described in detail immediately below; θ_j is a crop-specific fixed production cost that is independent of yield (excluding land rental costs, which is implicitly the outcome variable we are modeling); and rx is expenditure on the variable input x whose constant marginal cost is r.

Crop yields are given by the product of two terms: a generic function $f(\cdot)$, which maps fixed and variable inputs into a unitless measure of relative crop yields (i.e., yield divided by maximum potential yield), and a crop-specific constant μ_j, which scales this value to express crop yields in the proper units. Relative yields depend on a crop-specific index of fixed inputs z and variable inputs x. Fixed inputs are weighted by parameter λ_j, which differs for each crop according to the crop's specific needs. For example, if the crop has a long growing season, then λ_j will be large and positive to capture the importance of temperature in determining the crop's yield. We focus here on a single fixed input to simplify the notation and exposition, but the same result would obtain if we expanded the linear index to include multiple fixed inputs, each weighted by its own crop-specific parameter, as in our econometric analysis below.

Conditional on growing crop j, profit maximization with respect to the variable input then leads to profits as a function of prices, the fixed input, and production parameters:

$$\pi_j^* = p_j \mu_j f\left(f'^{-1}\left(\frac{r}{p_j \mu_j}\right)\right) + \theta_j - r f'^{-1}\left(\frac{r}{p_j \mu_j}\right) + r\lambda_j z, \tag{2}$$

where $f'^{-1}(\cdot)$ is the inverse first derivative of $f(\cdot)$ (see the derivation in the Appendix). The first term on the right-hand side is revenue, the second term is fixed costs, the third term is variable costs assuming a zero endowment of fixed inputs, and the fourth term is the monetary value of the fixed input endowment, which is linear in a crop-specific parameter $r\lambda_j$ and fixed inputs z. We implicitly assume that the farmer chooses a strictly positive level of variable inputs, which is true so long as marginal revenue product evaluated at the fixed input endowment exceeds marginal cost, that is, so long as $p_j \mu_j f'(\lambda_j z) > r$.

One key feature of this model is that changes in output prices shift profits up and down but do not alter the value of the fixed inputs, which are pinned down by variable input prices. A second, closely related feature is that yields are independent of fixed inputs, varying only with input and output prices. Intuitively, since variable inputs can be used to compensate for the lack of fixed inputs without diminishing returns, anticipated yields at the optimum are independent of the fixed input endowment. These features seem reasonable in some cases, for example in the case of fertilizer compensating for a soil's lack of nutrients or irrigation compensating for the lack of rainfall. These features could be problematic in the case of mean temperature, however, for which there is no close substitute, although one could imagine using a combination of fast-growing seed varieties and more intense use of other resources, such as investment in greenhouses. To relax this assumption would require a more sophisticated econometric model that is beyond the scope of this paper.

What does our model imply about the relative profitability of different crops across locations? If we take this model literally, then if two farmers face the same input and output prices and choose to grow the same crop, the first three terms in the profit function above (capturing revenues, fixed costs, and variable costs) will be identical; their overall profits will differ only to the extent that they face different fixed input endowments in the fourth term. Thus, by invoking an assumption of uniform prices, we are left with an expression for variable profits that is linear in a crop effect shared by all parcel owners (the first three terms) and a vector of soil and climate characteristics that differs across landowners (the fourth term). We lean on the structure of our conceptual model and the assumption of uniform prices in deriving our empirical model and in our policy simulations below.

In practice, output prices are fairly uniform across locations, since transportation costs for most agricultural commodities are relatively low. Prices do vary somewhat geographically, but much of this variation is common to all crops.[7] On the input side, prices for most variable inputs are also fairly uniform across locations, which is not surprising, given that many inputs, including chemical fertilizers and pesticides, are tradable commodities with relatively low transportation costs, while capital and agricultural labor are also highly mobile across locations.[8]

These arguments for uniform prices are less valid in the case of irrigation water. The availability of fresh surface water varies geographically, as does the presence and depth of freshwater aquifers, so that the long-run marginal cost of locally procured water may vary substantially. When irrigation water is available at low cost, two things happen in our model. First, farmers use more water in production, which increases yields and therefore revenues, while variable costs on infra-marginal irrigation water consumption fall, both of which increase variable profits. Second, and more problematic for our analysis, cheap irrigation water devalues annual rainfall as a fixed input endowment, which increases the relative profit of

[7]We collected state-level prices for 1996–1998 (the middle of our sample period) for two-dozen agricultural commodities from USDA NASS, available here: `http://usda.mannlib.cornell.edu/MannUsda/viewDocumentInfo.do?documentID=1050`. We then regressed logged prices on a vector of crop and state dummies. These controls explained 98% of the variation in prices in each of these years. We also examined the price data crop-by-crop. We found that the standard deviation in logged prices across states was roughly 0.1 or less for most crops, rising only as high as 0.25–0.30 for potatoes, sweet potatoes, and alfalfa.

[8]We collected data on region-level fuel, phosphate, and nitrogen prices from USDA NASS, available here: `http://quickstats.nass.usda.gov/`. The standard deviation in logged prices across regions in 2001 was roughly 0.1 or less. We also collected data on region-level wage rates for field workers in 1997 from USDA NASS, available here: `http://usda01.library.cornell.edu/usda/nass/FarmLabo//1990s/1997/FarmLabo-08-15-1997.txt`. The standard deviation in logged wages across regions was only 0.08.

planting a water-intensive crop. In short, while input prices are unlikely to be key determinants of a crop's relative profitability in most cases, this is probably not true for water. Thus, as a robustness check, we also estimate our model separately for Eastern states, where rainfall is relatively abundant and irrigation is much less common, as in Schlenker et al. (2006).

2.2 Crop choices

Using the conceptual model above as motivation, we now develop an empirical model of crop choice that we can estimate using historical data. Let the expected profit to individual parcel owner i living in county c that chooses to grow crop j be given by the following expression:

$$\pi^*_{icj} = \beta'_j z_c + \gamma_j + \delta_c + \xi_{cj} + \epsilon_{icj}, \tag{3}$$

where: $\beta'_j z_c$ captures the crop-specific value of the county's endowment of fixed inputs (the fourth term in equation 2), which is, as in the conceptual model above, linear in a vector of county soil and climate characteristics (z_c) and a crop-specific parameter vector ($\beta_j = r\lambda_j$) that reflects both variable input prices and the crop's specific needs; γ_j is a crop effect, assumed to be constant across counties, that captures differences across crops in variable profits and fixed production costs (the first, second, and third terms in equation 2); δ_c is a county effect, assumed to be constant across crops, that captures geographic variation in variable profits and fixed production costs; ξ_{cj} is a crop-specific profit shifter shared by all parcel owners in the same county; finally, ϵ_{icj} is a random error term that captures parcel-specific variation in the expected profits from growing a given crop, arising from unobserved variation in variable profits, fixed costs, and fixed input endowments.

To facilitate estimation, we assume that the random error term ϵ_{icj} follows an i.i.d. extreme value distribution with variance $\sigma^2\pi^2/6$ and that parcel owners choose whichever crop yields the highest expected profits. As is well known, these assumptions lead to the multinomial logit model and its convenient closed-form expression for the probability that an individual landowner chooses a given crop (see Train, 2009). Given that we model choices as a function of county-average characteristics, this probability equals the share of a county's available cropland devoted to a particular crop as a function of county averages:

$$s_{cj} = \frac{\exp((\beta_j' z_c + \gamma_j + \delta_c + \xi_{cj})/\sigma)}{\sum_{k=0}^{K} \exp((\beta_k' z_c + \gamma_k + \delta_c + \xi_{ck})/\sigma)}, \tag{4}$$

where the summation with respect to k in the denominator is over all K crops in the farmer's choice set. If not all crops in the choice set are included, then this equation gives crops shares conditional on the set of crops being modeled explicitly. We discuss the choice set in greater detail below.[9]

As is well known, dividing these shares by the share of a county's cropland devoted to a reference land-use category and then taking logs yields a model that is linear in parameters and that can therefore be estimated using conventional linear estimation techniques:

$$\ln s_{cj} - \ln s_{c0} = \tilde{\beta}_j' z_c + \tilde{\gamma}_j + \tilde{\delta}_c + \tilde{\xi}_{cj}, \tag{5}$$

where s_{c0} is the share of county c's available cropland devoted to the reference land-use category, say idle cropland or CRP land, and we have implicitly normalized expected profits for the reference category to zero.

We highlight two key features of this model. First, the estimated parameters $\tilde{\beta}_j = \beta_j/\sigma$, the estimated crop effects $\tilde{\gamma}_j = \gamma_j/\sigma$, the estimated county effects $\tilde{\delta}_c = \delta_c/\sigma$, and the estimated residuals $\tilde{\xi}_{cj} = \xi_{cj}/\sigma$ all have a structural interpretation in terms of expected profits (normalized by the scaling parameter σ, as denoted by the tildes) and can be used to predict crop shares using equation (4) above. Second, since $\ln s_{c0}$ is perfectly collinear with

[9] As demonstrated in Timmins (2006), if one makes a distributional assumption on the error term and assumes that landowners choose to grow the crop that maximizes land value, then there is an explicit link between land values and crop choices. That is, the hedonic land value and crop choice models are connected. With data on both land values and crop choices, this link then can be exploited to improve efficiency in estimation. We leave this approach as a potential avenue for future research.

our county fixed effects, estimates for the parameter vector $\tilde{\beta} = \beta_j/\sigma$ are invariant to our choice of reference category and are consistently estimated even if we ignore the reference category completely in our estimation; the choice of reference category only becomes relevant when we predict crop shares according to equation (4) above. We return to the issue of the reference category below.

One could imagine estimating the parameter vector $\tilde{\beta}_j = \beta_j/\sigma$ separately for each individual crop, and some studies indeed estimate crop-specific parameters; this would give our model the traditional multinomial logit interpretation. Since our main goal is to predict behavior for new crops that do not already exist in the historical data, however, we will attempt to explain why these production parameters vary across crops as a function of observable crop attributes. Thus, we assume that the vectors of crop-specific parameters take the following form:

$$\tilde{\beta}_j = \alpha_0 + \sum_{a \in A} \alpha_a w_{aj} \tag{6}$$

where w_{aj} is the ath attribute (among attribute set A) of crop j and α_a is a column vector of coefficients (many of which will be restricted to zero) capturing how the crop-specific production parameters depend on crop attributes. This assumption gives our model the traditional conditional logit interpretation. We can estimate the parameters in α_a (for $a \in A$) simply by interacting observable soil and climate characteristics with observable crop attributes.[10] For example, we can include an interaction between average annual rainfall and water-use efficiency to account for the fact that water-sensitive crops will tend to be more profitable in areas with higher expected rainfall, all else equal. In this way, the model explicitly ties the unique physiological attributes of each crop to the land and climate characteristics of different locations.[11]

[10] Note that the parameters in α_0 are not identified in our main specification, as the inner product of α_0 with the county characteristics z_c is the same for all crops in a given county and therefore perfectly collinear with our county fixed effects.

[11] In discrete-choice studies of demand for differentiated products, modelers typically specify utility as a linear function of product attributes, allowing the parameters on these attributes to vary predictably with a household's characteristics. Here, we take the reverse tack, specifying profits as a linear function of a parcel's soil and climate characteristics and allowing the parameters on these inputs to vary predictably with crop

As described above, we implicitly assume that expected output prices, variable input prices, and fixed costs vary negligibly across locations, so that our crop dummy variables capture variable profits and fixed costs given by the first three terms in our profit function. We assume that any remaining variation in profits is roughly uniform across all crops that might be grown in a given location and therefore captured by the county dummy variables.

2.3 Choice set and reference category

We define our choice set to include all field crops whose attributes we observe, which includes almost all field crops, a handful of vegetable crops whose attributes we observe, and all CRP land, which together account for 70% of U.S. cropland. See Table 1 for the full list of crops. We omit other agricultural land, including pasture and rangeland, as well as forested areas. Thus, our analysis gives land-use choices conditional on growing the above crops or enrolling in CRP. We leave the choice of cropland versus other land-use categories to future research. In defining the choice set to include all of these crops, we implicitly assume a long-run or steady-state model of crop supply, in which crop choices have fully adjusted to prices, climate, and technology.

We choose CRP land as our reference category, so that the county dummies in our model reflect expected profits for crops relative to CRP, which provides land rental payments to farmers in exchange for removing their cropland from production for a contracted period of years.[12] This choice makes sense for several reasons. First, CRP land accounts for a large

attributes. While the two approaches are mathematically equivalent, our specification flows more naturally from our conceptual model of crop production, in which the underlying parameters are marginal products of soil and climate characteristics that differ by crop.

[12]In practice, the CRP enrollment process approximates a market mechanism. See here for details on the program's administration: `http://www.fsa.usda.gov/FSA/webapp?area=home&subject=copr&topic=crp`. Farmers submit offer bids, which are ranked based on bid price and environmental benefits and then accepted or rejected subject to federal funding constraints and caps on county CRP acreage. Maximum payments are determined locally based on land rental rates and soil productivity; farmers may offer lower bids to increase the probability of acceptance. (This bidding process describes the General Signup program, which accounts for 88 percent of CRP acreage as of 2008. The Continuous Signup program, which accounts for 8 percent of acreage as of 2008, offers fixed, non-competitive payments for high-priority conservation areas and practices that are twice as those offered under the General Signup program. See Ferris and Siikamki (2009).) Given this process, fluctuations in CRP funding or acreage constraints can be viewed as shifts in

share of cropland and is adopted by a sizeable fraction of landowners in most counties.[13] Thus, we can usually calculate a crop's share relative to CRP directly without dividing by zero. Second, CRP land is an important margin upon which farmers can easily expand crop production (after their current contracts expire). Third, the return to choosing CRP is determined directly by government policy. This is relevant since in future work we plan to simulate the effects of climate change on cropping patterns. Measuring the profitability of other crops relative to CRP therefore makes sense conceptually, since climate change does not directly impact CRP's return.

2.4 Zero shares and distributional assumptions

Unfortunately, since the distributions of some crops are quite limited geographically, there are many cases of zero crop shares in our crop-county panel dataset. These zeros are problematic for two reasons. First, if we take our distributional assumption literally, such zero shares would be highly unlikely in a large county, since the iid extreme value distribution has infinite support. Second, these zero shares lead to missing values after taking logs, which raises the more practical issue of how to handle these missing values.

Our main approach is to add ten acres (or some other small number) to the acreage of each crop in each county before calculating crop shares, as in Timmins (2006). We explore the sensitivity of our results to adding different acreage values and to dropping zeros altogether. We also explore the sensitivity of our results to including crop-region fixed effects, estimating our model using only the 5–10 most widely grown crops, and estimating a variant of our model using a quasi-likelihood approach that can handle zero shares without resorting to such ad-hoc fixes (Papke and Wooldridge, 1996; Mullahy, 2010), all of which address the problem

demand for CRP land's output, namely environmental services, while fluctuations in CRP payments holding acres fixed are determined mainly by the return to growing other crops. Thus, while CRP land is special in some ways, it is appropriate to include it with the other crops in the choice set. Farmers commit to 10–15 year contracts when they enroll their land in CRP. Of course, farmers also make long-term commitments when they choose to grow perennial crops or invest in crop-specific skills or machinery. We implicitly assume that any option value related to uncertain returns is captured by our crop fixed effects.

[13] See here: `http://www.fsa.usda.gov/Internet/FSA_File/su41county.pdf`

of zero shares either directly or indirectly. Overall, our results are robust to these different approaches. Note that our inclusion of crop-region fixed effects is also a crude approach to relaxing the independence assumption for the idiosyncratic errors, as it captures correlation between unobserved crop attributes and county characteristics that happen to be spatially correlated. We also relax the independence assumption directly by estimating a nested logit model as part of our robustness checks.

3 Plant growth and crop production

In this section we discuss important conceptual issues in modeling plant growth and crop production. Our goal is to motivate our choice of soil and climate characteristics, as well as our choice of crop attributes, which we hypothesize to influence the relative profitability of different crops across locations. We discuss which variables should be included, as well as how these variables should be included, paying close attention to important functional form and specification issues.[14]

3.1 Temperature

Plant growth is sensitive to temperature. If a crop experiences many days of optimal temperatures, then its end-of-season yield will be high, while if the crop experiences many days that are either too hot or too cold, its end-of-season yield will be lower. Although previous research has carefully estimated the precise, functional relationship between daily temperatures, growth rates, and end-of-season yields for some crops, including for example corn, cotton, and soybeans (Schlenker and Roberts, 2009; Roberts and Schlenker, 2010b), this level of detail is unfortunately not available for most crops. Thus, in our empirical application, we interact various measures of temperature with proxies for a plant's specific temperature

[14]This section relies heavily on common knowledge about plant growth and crop production, which we have absorbed through discussions with colleagues in crop and soil science, as well as from a summer course in plant growth modeling taught by Bruno Basso at the W.K. Kellogg Biological Station in 2011. Any errors in translation or understanding are purely our own.

needs. Some crops are more tolerant of extreme cold, while other crops are more tolerant of extreme heat. We measure extreme cold as degree-days below 0C and extreme heat as degree days above 30C. We interact extreme cold with a plant's minimum tolerable temperature, which is a proxy for cold tolerance. We interact extreme heat with a plant's radiation-use efficiency (which we define precisely below) and an indicator for whether the plant uses C3 or C4 photosynthesis, since C4 crops tend to be more heat tolerant (Kiniry, Jones, O'toole, Blanchet, Cabelguenne, and Spanel, 1989). Plants also differ in the total, accumulated amount of heat energy they need to develop, mature, and produce high yields. We measure the level and duration of average temperatures within the optimal window for growing crops as degree-days between 0C and 30C. We interact this measure with a plant's minimum frost-free days, which is a proxy for the length of the plant's growing season.

3.2 Sunshine

During photosynthesis, plants convert solar energy and carbon dioxide into carbohydrates, which plants use to fuel their growth. Thus, areas with ample sunshine are typically more suitable for growing crops. But some plants are more sensitive to sunlight than others. One widely available proxy for the sensitivity of plant growth to sunlight is a plant's radiation-use efficiency, which quantifies how many units of carbohydrates per unit of solar energy the plant is able to generate. Yields for plants with high radiation-use efficiency tend to be more sensitive to sunlight. In addition, plants that rely on C4 photosynthesis use sunlight more efficiently than C3 plants, and tall plants may be able to capture more light. Thus, in our empirical application we interact measures of average sunshine with a plant's height, radiation-use efficiency, and an indicator for whether the plant uses C3 or C4 photosynthesis.

3.3 Water

All plants need water to grow. Thus, areas with more rainfall are typically better for growing crops, as are soils that retain more water, since rainfall and available water capacity interact to determine the amount of water available for plant growth.[15] One widely available proxy for the sensitivity of plant growth to water inputs is water-use efficiency, which is defined as total yield divided by total water consumption.[16] Crops with high water-use efficiency will come nearer to attaining their potential yield where rainfall is plentiful and the soil holds plenty of water. In addition, crops with long roots are better able to draw water from the soil and therefore less likely to suffer in dry conditions. Thus, in our empirical application we interact measures of annual rainfall and a soil's available water capacity with a plant's water-use efficiency and root depth. We expected these interactions to be less important in locations with access to irrigation.

3.4 Soil nutrients

A soil's chemical composition has important implications for plant growth. Nitrogen and other soil nutrients are critical for plant growth, so areas with more naturally occurring nutrients are more productive. Of course, fertilizer can be used to supplement in other areas. Soil that is too acidic or too basic can harm plant growth. Finally, soil that is too salty can also impair growth. These soil features will have differential impacts on production costs, depending on a crop's attributes. Some crops demand more nutrients, while other crops need less. Crops that grow well in high-pH soil may perform poorly in low-pH soil and vice-versa. Finally, some crops are tolerant of salty soil, while others are not. Thus, in our empirical application, we interact measures of a soil's nitrogen content, organic matter, salinity, and pH with a crop's ability to fix nitrogen, fertility needs, salt tolerance, and optimal pH range.

[15] A soil's water-holding capacity is the total amount of water that the soil will hold (inches of water suspended per inch of soil), while a soil's available water capacity is the fraction of this total that plants are able to draw from the soil to sustain their growth. We use the latter measure in our analysis.

[16] Assuming that yield as a function of water passes through the origin, then water-use efficiency is mathematically equivalent to average marginal yield with respect to water. Thus, water-use efficiency proxies for average sensitivity to water.

3.5 Soil conservation

Some soils are particularly susceptible to wind and water erosion, and the loss of this soil could damage future productivity, while other soils are depleted of nutrients and require further investment. Not all crops are created equal when it comes to maintaining soil quality. In areas that are highly susceptible to erosion, tall crops and crops with deep roots can help prevent the soil from being blown or washed away. Thus, we would expect such crops to be more desirable in areas that are susceptible to erosion. In addition, some crops, such as corn, deplete a soil's nitrogen content, while other crops, such as soybeans, add nitrogen back into the soil. Thus, we might expect the incentive to plant a nitrogen-fixing plant to be higher in nitrogen-depleted soils. In our empirical application, we interact measures of a soil's susceptibility to wind and water erosion with a crop's height and root depth, and we interact a soil's nitrogen content with an indicator for whether the crop fixes nitrogen.

3.6 Crop rotation

Crop rotation arises through dependence over time in the return to growing different crops on a particular parcel of land. In a long-run, steady state model such as ours, this amounts to farmers choosing, not individual crops as we have assumed, but rather bundles of crops that, when grown in rotation, yield higher average returns than any individual crop grown continuously. While we do not model crop rotation explicitly, some aspects of crop rotation are captured implicitly in our model, and we address other aspects of crop rotation in our robustness checks.

Crop rotation is used partly to maintain a soil's nitrogen content. As noted above, for example, corn depletes nitrogen, while soybeans fix nitrogen. Thus, if a farmer grows corn one year, her soil will be in a nitrogen-depleted state the next year, raising the future return to growing soybeans. In effect, the farmer is choosing a corn-soybean rotation rather than corn and soybeans individually. In ignoring crop rotation, we would therefore tend to associate corn's ideal growing conditions with soy's crop attributes, potentially leading to biased coefficient estimates. If the marginal value of adding nitrogen to the soil is roughly constant, however, then the added benefit of growing a nitrogen-fixing crop will tend to be captured by our crop dummies. Similarly, if fertilizer prices are low, as they were during much of our sample period, then this value will not be particularly large to begin with. In either case, ignoring crop rotation to maintain soil fertility is less problematic.

Crop rotation is also used to inhibit weeds, fungi, and other pests from establishing themselves in the soil. While we cannot model specific pairings of crops that are particularly profitable for this reason, a generic return to crop diversity would manifest empirically as a weaker correlation between observed crop attributes and county soil and climate characteristics, or equivalently, as a higher variance on the idiosyncratic error term. Again, if chemical pesticide prices are low, this incentive for crop rotation will not be particularly large to begin with.

We further address these incentives for crop rotation through three robustness checks. First, we interact a crop's level of nitrogen fixation with a county's soil nitrogen content, which is available for some counties. While this approach does not directly capture the choice of nitrogen-fixing crops to offset nitrogen-depleting crops, it does capture long-run variation across counties in naturally occurring nitrogen. Second, we estimate a nested logit model in which crops that are often grown together in rotation (or that are otherwise geographically correlated) share the same nest. Third, we estimate a model that includes crop-region fixed effects, which captures unobserved region characteristics that might lead particular crops or cropping rotations to be especially profitable.

3.7 Perennial crops

While most crops in our choice set are annual crops that must be planted anew each spring, our choice set also includes several perennial crops, including alfalfa and sugarcane. Perennial crops can be harvested repeatedly for several years without replanting, reducing planting costs. At the same time, perennial crops may be more costly to abandon once their root systems become established in the soil. For the most part, these unobserved attributes should be captured by our crop dummies. If the unique features of perennial crops interact strongly with soil, climate, or other location-specific characteristics, however, then our failure to model these interactions directly my lead to bias. For example, if farmers in locations susceptible to erosion plant perennial crops because they can leave the root systems intact all year, and we ignore this interaction in our model, then our coefficient estimates for other variables may be biased.

4 Data sources

In this section we describe our data sources for crop choices, crop attributes, and county soil and climate characteristics, as well as the construction of our key explanatory variables.

4.1 Crop choices

We obtain county-level annual data on planted acres for major crops for 1986–2008 from the USDA National Agricultural Statistics Service.[17] We obtain county-level annual data on acres of land enrolled in the CRP program during 1986–2008 from the USDA Farm Service Agency.[18] We calculate for each county the share of cropland held by each of these crops on average during 1986–2008. We choose 1986 as our first year, since the modern Conservation Reserve Program began in this year. We calculate average crop shares over the entire sample period, rather than crop shares in any particular year, to reduce statistical sampling errors, to minimize the occurrence of zero shares for some crops and counties, and most importantly to ensure that our cross-sectional results are capturing long-run behavior rather than short-run phenomena. Our empirical model of crop choice is based on expected profits. Thus, we calculate crop shares using planted acres, since harvested acres may be endogenous to

[17] Available online: `http://www.nass.usda.gov/Charts_and_Maps/Crops_County/Data/index.asp`
[18] Available online: `http://www.fsa.usda.gov/Internet/FSA_File/historycounty.xls`

Table 1: County crop shares (percents)

Crop	Mean	Std. Dev.	Minimum	Maximum
Alfalfa	8.91	16.19	0.00	94.91
Barley	1.92	4.50	0.00	45.63
Beans	0.43	1.77	0.00	28.91
Chickpeas	0.12	0.42	0.00	8.67
Corn	26.13	20.49	0.00	88.35
Cotton	4.11	10.28	0.00	70.70
Flaxseed	0.15	0.41	0.00	4.56
Mustard	0.11	0.30	0.00	4.24
Oats	4.29	8.33	0.00	96.06
Peanuts	1.13	4.67	0.00	62.35
Peas	0.16	0.60	0.00	13.65
Potatoes	0.62	4.00	0.00	92.65
Rapeseed	0.20	0.92	0.00	21.81
Rye	0.79	3.28	0.00	64.94
Safflower	0.11	0.31	0.00	4.24
Sorghum	3.68	7.32	0.00	81.30
Soybeans	19.10	17.92	0.00	81.30
Sugarbeet	0.35	1.50	0.00	28.28
Sugarcane	0.54	5.84	0.00	99.93
Sunflower	0.43	1.37	0.00	19.54
Tobacco	1.17	4.51	0.00	57.03
Tomatoes	0.18	1.31	0.00	31.40
Wheat	16.51	17.81	0.00	92.09
CRP	8.86	11.14	0.00	76.53

Note: Table reports summary statistics for average county crop shares during 1986–2008. Mean and standard deviation are for the un-weighted estimation sample of 2541 counties.

mid-season fluctuations in weather and crop prices.[19] Table 1 shows county average crop shares during 1986–2008 for the crops and counties in our estimation sample.

[19] Alfalfa, sugarcane, and tobacco are not always planted every year. Thus, we calculate their crop shares using harvested acres, which is a better measure of land devoted to these perennial crops.

Table 2: Crop attributes

Item	Frost-free days	Min. temp	Root depth (in)	WUE index	RUE (g/MJ)	Photo-synth.	Height (in)	Salt tol.	Min pH	Max pH	Fertility req.	Nitrogen fixation
Alfalfa	90	-43	24	0.43	1.10	C3	2.0	Medium	6.0	8.5	High	High
Barley	90	-43	10	0.85	1.60	C3	2.5	High	5.0	8.5	Medium	None
Beans	120	47	6	0.80	1.00	C3	3.0	Low	6.0	6.9	Medium	Medium
Chickpeas	115	-43	16	0.64	1.00	C3	3.0	Medium	6.0	8.5	Low	High
Corn	90	32	8	1.00	2.00	C4	8.0	Low	5.5	7.5	High	None
Cotton	365	50	16	0.65	1.50	C3	6.0	None	6.0	7.0	Medium	None
Flaxseed	100	-43	2	0.21	1.70	C3	2.6	None	5.0	8.0	Medium	None
Mustard	125	-13	12	0.32	1.92	C3	3.0	Medium	5.0	8.0	High	None
Oats	90	-23	8	0.33	1.45	C3	2.0	Medium	5.3	8.5	Medium	None
Peanuts	265	12	20	0.36	1.20	C3	1.3	Low	5.0	7.5	Medium	Medium
Peas	180	7	16	0.58	1.50	C3	1.6	Medium	5.8	7.8	Low	Medium
Potatoes	90	45	10	1.72	1.65	C3	2.0	Low	5.2	6.8	High	None
Rapeseed	130	17	6	0.29	1.40	C3	4.0	None	6.0	7.2	Medium	None
Rye	110	-33	8	0.35	2.00	C3	3.5	Medium	4.5	8.2	Medium	None
Safflower	120	-20	10	0.35	1.45	C3	3.0	Medium	6.0	7.0	Medium	None
Sorghum	90	47	12	1.12	1.75	C4	4.0	Medium	5.5	7.5	High	None
Soybeans	140	-28	8	0.65	1.00	C3	3.0	Medium	5.5	7.8	Low	Medium
Sugarbeet	90	-5	2	0.00	1.90	C4	2.0	High	6.5	7.0	High	None
Sugarcane	365	17	24	1.29	1.75	C4	12.0	Medium	4.0	7.0	Medium	None
Sunflower	80	52	8	0.48	1.85	C3	9.0	Medium	5.5	7.8	Low	None
Tobacco	120	65	24	0.84	1.52	C3	6.0	Medium	5.7	7.8	High	None
Tomatoes	365	32	24	1.10	1.00	C3	6.0	Medium	5.5	7.0	High	None
Wheat	100	-28	18	0.71	1.65	C3	3.3	Medium	5.5	8.0	Medium	None
Mean	149	4.4	12.7	0.66	1.52		4.0		5.5	7.6		
Std. Dev.	94	36.6	7.0	0.40	0.33		2.7		0.6	0.6		
Minimum	80	-43	2.0	0.00	1.00		1.3		4.0	6.8		
Maximum	365	65	24.0	1.72	2.00		12.0		6.5	8.5		
Range	285	108	22.0	1.71	1.00		10.7		2.5	1.7		
Miscanthus	120	-18	14	1.18	2.00	C4	5.0	None	4.6	7.5	Medium	None
Switchgrass	120	-43	12	0.94	2.20	C4	5.0	Medium	4.5	8.0	High	None

4.2 Crop attributes

Table 2 presents attributes for the roughly two-dozen crops that we include in our analysis. We obtain these data from multiple sources. We obtain data on radiation-use efficiency (RUE) from several published papers in the crop and soil science literature, as described in the Appendix. Plants that use C4 photosynthesis tend to have higher RUE than C3 plants (Kiniry et al., 1989). Thus, some of our proxies for heat tolerance will be correlated with each other.

We obtain data on water-use efficiency (WUE) from roughly a dozen sources in the crop and soil science literature, all documented in the Appendix. Several different metrics are used to measure water-use efficiency in the literature. These metrics are all based on average plant growth, measured in terms of biomass, CO2 intake, or grain mass, per unit of water vapor loss. We use the biomass measure, as it is the best measure of overall plant growth for comparison across crops.[20] While the studies we survey collectively report WUE for all crops, no individual study covers all crops. Thus, we construct a crop-specific index of WUE using a regression-based approach that controls for potential differences across studies.[21]

Finally, we obtain data on various other crop attributes from the USDA Natural Resources Conservation Service's PLANTS Database, which aggregates information from the scientific literature, agency documents, and the accumulated knowledge of plant specialists.[22] Attributes include a plant's height, root depth, minimum frost-free days (a proxy for length of growing season), and minimum and maximum pH levels, as well as qualitative measures of the plant's shade tolerance, drought tolerance, salinity tolerance, calcium tolerance, soil fertility requirements, and ability to restore a soil's nitrogen content (nitrogen fixation). The

[20] From the standpoint of our conceptual model, a superior measure would be marginal logged yield with respect to water inputs, which targets marketable yield directly yet is independent of the units in which yield is measured. We are actively working to obtain this measure.

[21] We treat the reported estimates of WUE in these studies as a panel dataset. We then regress logged WUE on a set of crop dummies and study dummies. The exponential function evaluated at the coefficients on the crop dummies yields our index. Given our inclusion of study dummies, only studies that report water-use efficiency for two or more crops contribute to identification of the index.

[22] Available online here: `http://plants.usda.gov/java/`

USDA's database is missing attributes for a handful of our crops, including potatoes, safflower, sugar beet, and tobacco. We fill in attributes for missing crops based on information from a variety of published and unpublished sources.

Note that both switchgrass and miscanthus fall entirely within the range of attribute values spanned by the other existing crops, with the exception that switchgrass has a somewhat high radiation-use efficiency. Thus, our forecasts for the adoption of these new bioenergy crops below will, for the most part, represent "within-sample" forecasts in terms of observable crop attributes.

4.3 Climate characteristics

We obtain data on soil and climate characteristics from a variety of sources. Table 3 presents summary statistics for these variables. Our measures of temperature and precipitation are based on historical daily temperatures and precipitation on a 4km-by-4km grid covering the entire United States from Schlenker and Roberts (2009).[23] We use these data to calculate, for each county, average monthly degree days above 30C, average monthly degree days within 0C–30C, and average monthly degree days below 0C for the months of March–November based on historically observed temperatures during 1950–2005.[24] These same data contain monthly precipitation (cm) by county, which we use to calculate average precipitation for the months of March–November during 1950–2005. We obtain county-level data on average

[23]Schlenker and Roberts (2009) generate these data through a combination of Parameter-Elevation Regressions on Independent Slopes Model (PRISM) data, which measure monthly temperatures and precipitation on a 4km-by-4km grid, and National Climactic Data Center (NCDC) data, which measure daily temperatures and precipitation at individual weather stations scattered throughout the country. To get daily temperatures and precipitation on the 4km-by-4km grid, they first regress the monthly measurements from the PRISM data on *monthly* averages for nearby weather stations from the NCDC data, and then apply their coefficient estimates to the *daily* measurements from the NCDC data to get predicted daily values at the grid points. They perform an out-of-sample validation exercise showing that their predictions are highly accurate. Note that the original PRISM data themselves are based on actual measurements from thousands of weather stations throughout the country, which have been extrapolated to the detailed grid using a sophisticated interpolation model.

[24]Following Schlenker and Roberts (2009), we approximate degree days below 0C, above 30C, and between 0C–30C on each day in each 4km-by-4km grid point based on the daily low and daily high temperatures. We then take the average across all grid points in a county, for all days during March-November, for all years during 1950–2005.

sunshine (the percent of all daylight hours for which the sun is not obscured by clouds) from Albouy, Graf, Kellogg, and Wolff (2011).[25]

4.4 Soil characteristics

We obtain data on soil available water capacity (inches of water available for plant growth per inch of soil), pH, organic matter (percent concentration), nitrogen content (percent ammonium ion concentration), water erosion factor (susceptibility of soil particles to detachment and movement by water), wind erosion factor (tons per acre per year that could be lost to wind erosion), and salinity (percent concentration, as measured by electrical conductivity) from USDA's Soil Survey Geographic (SSURGO) Database.[26] These data measure soil properties across thousands of sub-county areas that share similar soil properties. For each of these sub-county areas, the data report soil properties for multiple strata of soil depth. We first calculate average soil properties for each sub-county area, weighting each soil stratum equally. We then aggregate to the county level, weighting by each sub-county area's share of total county acreage. We have complete data on the soil attributes listed above for approximately 2,860 of 3,000 counties nationwide. Unfortunately, the data for other counties are missing either because the survey lacks information on one or more soil attributes or because the data are not yet available in electronic format. In some cases, soil areas cannot be assigned uniquely to specific counties. This latter problem mainly applies to remote, agriculturally unproductive areas, such as deserts and national parks; the majority of agricultural areas are included in our analysis. We prefer SSURGO to other soil datasets that potentially have better geographic coverage, such as the National Resources Inventory (NRI), since SSURGO reports pH, available water capacity, and other important soil attributes not contained in these other datasets.

[25] These data are based on NCDC data for the 156 weather stations that record sunshine information. These data take the form of average sunshine by month-of-year. The authors calculate county-level data on sunshine by month-of-year via interpolation from the four closest weather stations to each county.

[26] Available online through the USDA's Soil Data Access website: `http://sdmdataaccess.nrcs.usda.gov/`

Table 3: County soil and climate characteristics

Variable	Obs.	Mean	Std. Dev.	Min.	Max.
Climate					
Degree days below 0C (DD0C)	2541	8.39	9.80	0.00	52.53
Degree days above 30C (DD30C)	2541	5.32	5.48	0.00	66.38
Degree days 0C−30C (DD0C−30C)	2541	497.84	102.76	229.76	761.06
Minimum temperature	2541	9.44	3.91	-2.76	19.96
Precipitation (cm)	2541	8.21	2.64	0.44	14.98
Sunshine (percent)	2541	62.92	4.97	41.41	88.15
Soil					
Available water capacity (in/in) (AWC)	2541	0.15	0.02	0.06	0.22
pH	2541	6.47	0.92	4.39	8.69
Organic matter (%)	2541	1.21	1.70	0.12	46.61
Ammonium content (%) (NH4)	1405	17.53	23.84	0.44	105.86
Water erosion factor	2541	0.32	0.06	0.11	0.55
Wind erosion factor	2541	65.82	25.21	0.00	250.00
Salinity (%)	2541	0.40	0.83	0.00	12.48

Note: Table reports summary statistics for county-level data on soil and climate characteristics. See text for details.

5 Econometric estimation and results

In this section we describe our econometric estimation of equation (5) above.

5.1 Identification and interpretation

Estimation of equation (5) raises several important issues of identification and interpretation. A key distinction in the crop supply literature is between models estimated primarily using cross-sectional variation in soil and climate characteristics (Mendelsohn et al., 1994; Schlenker et al., 2006; Massetti and Mendelsohn, 2011) versus those estimated primarily using short-run fluctuations in weather or prices (Deschnes and Greenstone, 2007; Fisher, Hanemann, Roberts, and Schlenker, 2011). Since we are primarily interested in forecasting the long-run supply of new bioenergy crops in response to policies that permanently shift demand, we rely on cross-sectional variation to identify the parameters of our model. As in other cross-sectional models, one worry is that we have omitted important spatial variables that affect the relative profitability of particular crops or their attributes, potentially biasing our coefficient estimates. To explore the sensitivity of our results to such variables, we include crop-region fixed effects in some specifications. We emphasize, however, that omitted variables bias is less of a concern here than in many other papers, given our focus on forecasting the adoption of new crops, rather than on estimating the damages associated with climate change; we mainly seek unbiased (and low-variance) predictions, not necessarily all-else-equal coefficient estimates.

To facilitate interpretation of our coefficient estimates, we normalize our continuous crop attributes prior to estimation by taking the difference of each variable from its mean and then dividing by its range. Thus, a one-unit increase in each of these variables is equivalent to moving from the crop with the smallest attribute value in the choice set to the crop with the largest value. We similarly normalize our continuous county characteristics by taking the difference of each variable from its mean and then dividing by its standard deviation. Thus, a one-unit increase in each of these variables is equivalent to a one standard-deviation increase in the county characteristic. These normalizations allow us to interpret most of our coefficients directly as the marginal effect of a standard-deviation change in a county soil or climate characteristic on the relative shares of crops with the smallest and largest attribute values (all else equal). We normalize crop attributes and county characteristics prior to calculating any interactions, which facilitates interpretation for county characteristics that enter quadratically or that are interacted with other county characteristics. Given our normalizations, these higher-order effects drop out at the mean, allowing us to interpret the coefficients on the linear terms directly as marginal effects evaluated at county mean values.

5.2 Estimation results

Table 4 presents our main estimation results. We will begin by discussing the baseline regression results in column (1), which deals with potential zero crop shares by adding 10

Table 4: Estimation results

Variable	(1) Base	(2) N-fix	(3) Region	(4) Top-10	(5) Drop 0s	(6) QMLE	(7) Nest	(8) East
DD 0C−30C × Season	1.133	1.059	1.025	2.641	1.923	2.709	0.174	1.094
	(0.029)	(0.037)	(0.044)	(0.068)	(0.162)	(0.131)	(0.013)	(0.040)
DD>30C × C3	-0.234	-0.143	-0.080	-0.366	0.154	0.202	-0.049	0.258
	(0.052)	(0.067)	(0.051)	(0.081)	(0.063)	(0.077)	(0.010)	(0.120)
DD>30C × RUE	0.084	0.180	0.154	-0.068	1.008	0.750	-0.023	0.324
	(0.036)	(0.057)	(0.036)	(0.138)	(0.123)	(0.128)	(0.008)	(0.100)
DD<0C × Min Temp	-0.709	-0.739	-0.571	-1.415	-0.188	-0.164	-0.131	-0.776
	(0.025)	(0.041)	(0.042)	(0.048)	(0.045)	(0.045)	(0.010)	(0.066)
Sun × C3	-0.157	-0.252	-0.345	-0.537	0.020	0.039	-0.063	-0.695
	(0.043)	(0.061)	(0.047)	(0.075)	(0.118)	(0.069)	(0.008)	(0.158)
Sun^2 × C3	0.178	0.148	0.121	0.273	0.122	-0.019	0.036	0.007
	(0.017)	(0.020)	(0.017)	(0.028)	(0.041)	(0.031)	(0.004)	(0.071)
Sun × RUE	0.500	0.309	0.071	2.400	0.432	1.182	0.065	0.195
	(0.039)	(0.051)	(0.039)	(0.158)	(0.134)	(0.129)	(0.009)	(0.129)
Sun^2 × RUE	0.043	0.037	0.019	0.553	-0.042	-0.254	0.016	0.130
	(0.016)	(0.025)	(0.017)	(0.075)	(0.052)	(0.057)	(0.003)	(0.058)
Sun × Height	-0.738	-0.801	-0.605	-3.889	-1.625	-2.004	-0.113	-0.554
	(0.033)	(0.046)	(0.048)	(0.157)	(0.226)	(0.169)	(0.011)	(0.084)
Sun^2 × Height	0.097	0.073	0.126	0.054	0.423	0.354	0.039	0.029
	(0.017)	(0.023)	(0.021)	(0.086)	(0.095)	(0.100)	(0.005)	(0.057)
Rain × Roots	-0.295	-0.209	-0.113	-0.411	-0.413	-0.301	-0.044	0.802
	(0.037)	(0.057)	(0.052)	(0.067)	(0.076)	(0.075)	(0.008)	(0.199)
Rain^2 × Roots	0.275	0.252	0.069	1.135	0.335	0.470	0.051	-0.488
	(0.032)	(0.047)	(0.037)	(0.063)	(0.052)	(0.065)	(0.007)	(0.114)
Rain × WUE	-0.113	-0.262	-0.156	1.302	-0.845	0.586	0.015	-1.852
	(0.045)	(0.069)	(0.061)	(0.141)	(0.113)	(0.154)	(0.009)	(0.218)
Rain^2 × WUE	-0.051	-0.067	0.082	0.412	0.096	0.254	-0.013	0.980
	(0.035)	(0.049)	(0.040)	(0.099)	(0.066)	(0.116)	(0.007)	(0.141)
AWC × Roots	-0.238	-0.144	-0.086	-0.179	-0.405	-0.371	-0.052	-0.026
	(0.032)	(0.051)	(0.032)	(0.063)	(0.058)	(0.066)	(0.007)	(0.067)
AWC^2 × Roots	0.047	0.052	0.058	-0.070	-0.085	0.005	0.010	0.092
	(0.017)	(0.029)	(0.017)	(0.034)	(0.040)	(0.047)	(0.004)	(0.023)
AWC × WUE	0.097	0.092	0.049	0.686	0.396	0.184	0.026	-0.095
	(0.035)	(0.056)	(0.036)	(0.099)	(0.078)	(0.100)	(0.007)	(0.068)
AWC^2 × WUE	0.084	0.107	0.019	-0.125	0.190	0.439	0.008	-0.011
	(0.022)	(0.033)	(0.023)	(0.065)	(0.059)	(0.067)	(0.005)	(0.027)
Rain × AWC × Roots	0.257	-0.035	0.195	-0.065	-0.126	0.009	0.057	-0.051
	(0.048)	(0.076)	(0.041)	(0.084)	(0.072)	(0.093)	(0.010)	(0.096)
Rain × AWC × WUE	0.098	0.347	-0.028	-0.250	-0.108	-0.099	-0.002	0.372
	(0.043)	(0.068)	(0.042)	(0.113)	(0.089)	(0.141)	(0.009)	(0.090)
pH × 1[pH<Min pH]	0.159	0.204	0.087	0.207	0.578	-0.017	0.034	0.117
	(0.028)	(0.034)	(0.028)	(0.076)	(0.068)	(0.064)	(0.006)	(0.033)
pH × 1[pH>Max pH]	-0.227	-0.135	0.102	-0.062	-0.114	-0.212	-0.019	0.444
	(0.033)	(0.061)	(0.033)	(0.070)	(0.059)	(0.056)	(0.007)	(0.108)
Organic × High Fertility	-0.003	-0.016	-0.017	-0.081	-0.219	0.014	0.005	0.028
	(0.021)	(0.025)	(0.015)	(0.073)	(0.065)	(0.042)	(0.003)	(0.015)
Organic × Medium Fertility	0.002	-0.002	-0.070	-0.243	-0.070	-0.086	-0.003	-0.000
	(0.018)	(0.018)	(0.014)	(0.074)	(0.046)	(0.059)	(0.003)	(0.018)
Salinity × Low Tolerance	-0.407	-0.493	-0.284	-1.157	-0.349	-0.292	-0.056	-0.104
	(0.046)	(0.086)	(0.039)	(0.118)	(0.081)	(0.148)	(0.008)	(0.087)

Salinity × Medium Tolerance	-0.373	-0.160	-0.239	-0.915	-0.136	-0.170	-0.066	-0.028
	(0.043)	(0.046)	(0.037)	(0.103)	(0.081)	(0.154)	(0.008)	(0.042)
Salinity × High Tolerance	0.130	0.143	-0.115	-0.111	0.345	0.094	0.015	0.035
	(0.054)	(0.132)	(0.050)	(0.100)	(0.122)	(0.173)	(0.010)	(0.092)
H20 Erosion × Roots	-0.030	-0.193	-0.045	-0.461	-0.075	-0.221	-0.006	-0.116
	(0.032)	(0.043)	(0.027)	(0.069)	(0.065)	(0.066)	(0.006)	(0.052)
Wind Erosion × Roots	-0.441	-0.496	-0.369	-0.807	-0.237	-0.671	-0.089	-0.517
	(0.031)	(0.037)	(0.031)	(0.083)	(0.072)	(0.081)	(0.008)	(0.033)
Rain × H2O Erode × Roots	-0.199	-0.053	-0.155	-0.078	-0.269	-0.146	-0.043	-0.121
	(0.043)	(0.063)	(0.035)	(0.075)	(0.068)	(0.077)	(0.009)	(0.072)
NH4 × High N Fixation		-0.293						
		(0.035)						
NH4 × Medium N Fixation		-0.072						
		(0.028)						
Within-group correlation ρ							0.792	
							(0.204)	
Observations	58,443	32,315	58,443	25,350	14,600	58443	58,443	30,107
Number of counties	2,541	1,405	2,541	2,535	2,353		2,541	1,309
R-squared	.076	0.100	.033	.248	.101		-	0.063

Note: The table presents estimation results for regression of crop shares on county characteristics interacted with crop attributes. Regression (1) is our base specification that adds 10 acres prior to calculating county crop shares and taking logs. Regression (2) adds interactions of a soil's NH4 content with a crop's ability to fix nitrogen. Regression (3) adds crop-region fixed effects. Regression (4) estimates the model for the top-10 crops in our sample. Regression (5) drops zero shares altogether. Regression (6) estimates a variant of the model using a quasi-likelihood approach. Regression (7) estimates a nested logit model using 2SLS; the estimated within-group correlation in errors (ρ) is given toward the bottom of the table. Regression (8) estimates the base specification for eastern states. Dependent variable in all regressions (except 6) is logged crop share divided by CRP share, and all regressions (except 6) include both crop and county fixed effects; the dependent variable in regression (6) is crop shares in levels, and this regression controls for county characteristics directly in lieu of county fixed effects. Standard errors in parentheses are clustered by county in all cases. R-squared presents fraction of variation explained after removing crop, county, and crop-region fixed effects (in the case of regression 3). See text for details.

acres to each crop's county acreage prior to calculating county crop shares.[27] We then discuss various robustness checks below in the following subsection.

Consider first the effects of temperature on relative crop shares. The first coefficient in column (1) implies that a 0.1 standard deviation increase in degree days between 0C–30C (the range most suitable for growing crops) leads to a 11.3% increase in the shares of crops that require the most frost-free days (a proxy for the length of a crop's growing season) relative to crops that require the fewest frost-free days—that is, the ratio of their

[27] There were almost no sign changes across regressions that added 0.1, 1, 10, and 100 acres prior to calculating county crop shares. In the few cases where signs differed, the coefficients were statistically insignificant in both cases. While the signs were quite robust across specifications, the magnitudes of the coefficients were moderately sensitive to adding different acreage values.

crop shares increases by 11.3%. Most of the other coefficients can be interpreted similarly. A 0.1 standard deviation increase in degree days above 30C (extreme heat) leads to a 2.34% decrease in the share of C3 crops relative to C4 crops (which are more heat tolerant) and a 0.84% increase in the shares of crops that that have the highest radiation-use efficiency (whose yields are more sensitive to sunlight). A 0.1 standard deviation increase in degree days below 0 (that is, an increase in extreme cold) leads to a 7.09% decrease in the shares of crops that are least tolerant of cold. All of these coefficients have the correct signs, reasonable magnitudes, and are statistically significant.

Now consider the effects of sunlight on relative crop shares. A 0.1 standard deviation increase in sunlight leads to a 1.57% decrease in the shares of C3 crops relative to C4 crops, a 5.00% increase in the shares of crops with high radiation-use efficiency (whose yields are more sensitive to sunlight), and a 7.38% decrease in the shares of the tallest crops. All of these coefficients, which reflect marginal effects evaluated at mean county characteristics, have the expected signs and are statistically significant. The coefficients on the corresponding quadratic terms indicate that the negative interactions with C3 and height diminish in size as sunshine increases, while the positive interaction with radiation-use efficiency increases with sunshine.

Now consider the effects of precipitation on relative crop shares. A 0.1 standard deviation increase in precipitation leads to a 2.95% decrease in the shares of crops with the longest roots and a 1.13% decrease in the shares of crops with the highest water-use efficiency (whose yields are more sensitive to rainfall). Both coefficients, which reflect marginal effects evaluated at mean county characteristics, are statistically significant, but the coefficient for water-use efficiency has the unexpected sign. The coefficients on the corresponding quadratic terms indicate that the negative interaction with root length diminishes as rainfall increases, while the negative interaction with water-use efficiency increases in magnitude.

Now consider the effects of a soil's available water capacity on relative crop shares. A 0.1 standard deviation increase in available water capacity leads to a 2.38% decrease in the shares of crops with the longest roots and a 0.97% increase in the shares of crops with the highest water-use efficiency (whose yields are more sensitive to water). Both coefficients, which reflect marginal effects evaluated at mean county characteristics, have the correct signs and are statistically significant. The coefficients on the corresponding quadratic terms indicate that the negative interaction with root length diminishes as available water capacity increases, while the positive interaction with water-use efficiency increases with available water capacity. The next two coefficients capture interactions between rainfall and available soil water capacity and their relationship to a plant's root length and water-use efficiency. While the coefficient involving water-use efficiency has the correct sign, indicating that water-sensitive crops are more likely to be planted were rainfall and available water capacity are simultaneously high, the coefficient involving long roots has the incorrect sign.

Now consider the effects of soil pH, organic matter, and salinity on relative crop shares. A 0.1 standard deviation increase in pH leads to a 1.59% increase in the shares of crops for which the soil pH is currently too low and a 2.27% decrease in the shares of crops for which the soil pH is currently too high, both relative to crops for which the soil pH is currently within the acceptable range. These coefficients have the expected signs and are statistically significant. A 0.1 standard deviation increase in organic matter leads to a 0.02% increase and a 0.03% decrease in the shares of crops with medium and high fertility needs, both relative to crops with low fertility needs. Both coefficients are small in magnitude and statistically insignificant. Finally, a 0.1 standard deviation increase in salinity leads to 4.07% decrease, a 3.73% decrease, and a 1.30% increase in the shares of crops with low, medium, and high salt tolerance, all relative to crops with zero salt tolerance. While the first two of these coefficients have the incorrect signs, there are only three crops in the omitted category with zero salt tolerance. Moreover, relative to crops with low salt tolerance, the shares of crops with medium and high salt tolerance correlate as expected with salinity.

Finally, consider the effects of erosion on relative crop shares. A 0.1 standard deviation increase in a soil's susceptibility to water erosion leads to a 0.30% decrease in the shares of crops with the longest roots, while a 0.1 standard deviation in susceptibility to wind erosion leads to a 4.41% decrease, both relative to crops with the shortest roots. Both coefficients have the unexpected signs, the former of which is statistically insignificant, and the latter of which is exacerbated in the presence of high rainfall.

Thus, our initial conclusion from the base regression results in column (1) is that the attributes of chosen crops correlate as expected with soil and climate characteristics, with some exceptions. The model appears to be performing fairly well overall.

5.3 Robustness analysis

In this section, we explore the sensitivity of our results to different modeling choices. Overall, the attributes of chosen crops continue to correlate as expected with soil and climate characteristics. The parameter estimates are somewhat sensitive to if and how we deal with zero crop shares, as well as to our inclusion of crop-region controls, but not dramatically so.

5.3.1 Soil nitrogen content

The regression in column (2) adds soil nitrogen content interacted with crop attributes at the cost of dropping nearly half of the counties in our original sample. A 0.1 increase in soil nitrogen content leads to a 2.93% decrease and a 0.72% decrease in shares of crops with high and medium levels of nitrogen fixation, both relative to crops with low levels of nitrogen fixation. Both coefficients have the expected signs and are statistically significant. Importantly, the coefficients on the other variables do not change drastically when we add these new variables to our model, and these other estimated coefficients change little when we limit the sample to counties for which we have complete data on all variables.

5.3.2 Crop-region fixed effects

The regression in column (3) adds crop-region fixed effects, which control for unobserved soil and climate attributes that may lead particular crops or their attributes to be especially profitable or unprofitable in particular regions.[28] In addition, since zero shares in our sample are highly spatially correlated, these controls mitigate potential bias related to our ad-hoc fix for the zero shares problem. When we include these controls, however, more than four-fifths of our coefficient estimates retain the same sign, with roughly similar magnitudes, and all but two of the coefficients that flip sign are statistically insignificant in one or both regressions.

5.3.3 Popular crops only

The regression in column (4) estimates our model using only the ten most popular crops, which constitute nearly 90% of the cropland in our sample. While focusing on popular crops mitigates bias related to our ad-hoc fix of the zero shares problem, omitting less popular crops throws away potentially useful variation in crop attributes, which may weaken our forecasts for new crops. More than two-thirds of coefficients retain the same signs, however, while roughly half of the coefficients that do change sign are statistically insignificant after dropping unpopular crops. The most noticeable change is that many coefficients increase in magnitude. This is not surprising, since including unpopular crops (often at an assumed level of 10 acres) tends to increase the sample variation in logged crop shares, pushing the coefficient estimates toward zero.[29]

5.3.4 Dropping zero shares

The regression in column (5) drops zero crop shares from the dataset prior to estimation, implicitly assuming that the dropped crops are excluded from the choice sets of the affected counties. While this assumption is clearly inappropriate, nearly three-quarters of our coefficient estimates retain the same signs after dropping zeros, and most of those that do flip sign are statistically insignificant after dropping zeros. These qualitative comparisons are nearly

[28]We use the USDA ERS's nine farm resource regions, which tend to share similar soils, climates, and crop patterns. For further information, see here: `http://www.ers.usda.gov/Publications/aib760/`.

[29]We also estimate our model using the five most popular crops, which constitute nearly 80% of cropland in our sample. Roughly half of the coefficients flip sign relative to our base regression. Thus, we suspect that a choice set of five crops is simply not sufficient to generate robust results for a model based on crop attributes.

identical when based on regressions that add values other than 10 acres prior to calculating county crop shares.

5.3.5 Quasi-likelihood estimation

Following Papke and Wooldridge (1996) and Mullahy (2010), the regression in column (6) estimates a slightly modified version of our empirical model using a quasi-likelihood approach, which can handle the 0s in our crop shares data without adding small numbers or resorting to other ad-hoc fixes.[30] Three-quarters of the coefficients retain the same signs after we apply this quasi-likelihood approach, and many have similar magnitudes. Thus, we conclude on the basis of this regression and those above that our results are robust overall to our ad-hoc fix for zero shares.

5.3.6 Nested logit model

The regression in column (7) estimates our model using a nested logit specification, which allows the idiosyncratic errors to be correlated for crops in the same researcher-defined groups, while maintaining the independence assumption for crops in different groups (see Train, 2009). See the Appendix for a detailed derivation of our model and estimation approach. We group crops based on common crop rotations and by inspecting empirical correlations in county crop shares.[31] In the spirit of Berry (1994), we estimate this model by including,

[30] Mechanically, we maximize a conditional logit log-likelihood function applied to our county-level data, in which the 1s and 0s that would normally appear in this logit model have been replaced by our fractional crop shares, yielding a quasi maximum likelihood estimator (QMLE) of our parameters. Papke and Wooldridge (1996) and Mullahy (2010) show that this approach leads to consistent parameter estimates under the identifying assumption that expected crop shares follow the multivariate logistic functional form (see equation 3). While our structural model above implies additive errors (the ξs) after taking logs, this alternative estimator implicitly assumes additive errors in levels. Thus, the main estimation approach above aligns more closely with our underlying structural model, on which we rely to simulate the addition of a new crop to the choice set. While the county effects above (the δs) are absorbed through fixed effects estimation, the nonlinear estimator here precludes such an approach. Thus, we model the county effect as a linear function of the county averages of our explanatory variables (Mundlak, 1978; Chamberlain, 1980), which in our application is (with one minor exception) the same as controlling directly for observed county characteristics. See the Appendix for further details.

[31] Our nests include (1) alfalfa, barley, oats, and potatoes, (2) beans, chickpeas, sugar beets, and tomatoes, (3) corn, soybeans, and tobacco, (4) cotton, peanuts, sorghum, and rye, (5) flaxseed, peas, and rapeseed, (6) mustard and safflower, and (7) sunflowers and wheat, and (8) sugarcane by itself.

for each crop, the logged county share of the crop's group as an additional right-hand-side explanatory variable in equation (5). The coefficient on this extra variable is $\rho/(1-\rho)$, where ρ is the correlation in idiosyncratic error terms for crops in the same group, which we assume to be the same across groups. Since this extra variable is mechanically correlated with the error term in our regression, we instrument for actual logged group share with the log of *predicted* group share calculated using the baseline regression results in column (1).[32] The inclusion of this extra variable implicitly scales the other coefficients by $1/(1-\rho)$. Thus, we multiply all of our raw coefficient estimates by $1-\rho$, calculate robust standard errors using the delta method, and report the results column (7) to facilitate comparison to the other regressions. Based on this approach, we estimate $\rho = 0.792$, which is statistically different from zero. These results indicate a strong geographic correlation in unobserved profitability for crops in the same group. We plan to explore this issue in depth in future work.

5.3.7 Parameter stability

Finally, the regression in column (8) estimates our model separately for states bordering the western bank of the Mississippi River (i.e., Minnesota south to Louisiana) and eastward, where annual rainfall is more plentiful and irrigation is less common. This approach roughly approximates that in Schlenker et al. (2005), who estimate separate hedonic land value equations for counties with more or less than 20% of harvested cropland under irrigation, and Schlenker et al. (2006), who estimate a hedonic model for agricultural land east of the 100th meridian. Overall, two-thirds of the coefficients retain the same signs as in the

[32]We calculate predicted shares according to the shares formula in equation (4) using our estimated crop-county interactions, crop dummies, and county dummies from regression (1), deliberately omitting our estimated residuals (the ξs) from this calculation. In effect, our instrument is the weighted average of the group's crop attributes interacted with county characteristics, with the weights given by the coefficients in our baseline regression. Since the estimated coefficients in the baseline regression are functions of county crop shares, the weights and therefore our instruments are technically endogenous. We suspect any bias is exceedingly low, however, since the estimated coefficients reflect average crop shares across nearly 3,000 counties. Moreover, constructing our instrument in this way allows us to concentrate power in a single instrument, avoiding potential pitfalls associated with including multiple, potentially weak instruments. Indeed, interactions between crop attributes and county characteristics will only receive a large weight if they are important determinants of crop choices.

baseline specification. Over half of the coefficients on the water-related variables flip sign, however, and the magnitudes for the other coefficients also change. These changes could reflect heterogeneity in parameter values based on the availability of irrigation, or imprecision related to a smaller sample size and narrower variation in climate characteristics, or both. In any case, in future work we plan to address the availability of groundwater resources explicitly.

We also estimated the model separately for average crop shares during 1986–1999 and 2000–2008. This approach addresses the potential concern that crop choices depend on expectations about future prices, policy, climate, technology, and those expectations may have shifted significantly in the last decade due to rising global demand for food and energy, an increase in extreme weather events that foretell climate changes, and government policies to promote corn ethanol production. We find little difference in estimated coefficients on crop-county interactions across the two time periods, however, which gives us greater confidence in the use of our model to forecast near-future behavior based on observed crop choices in the recent past.

5.4 Model validation

While sensible coefficients give us confidence that our model is working well, our main goal is to forecast the adoption of new crops. Thus, in this section, we assess our model's ability to forecast where new crops will be grown by evaluating the quality of our model's out-of-sample predictions. Figure 1 presents a wide array of scatter diagrams. Each diagram plots, for a particular pairing of crops, the actual logged ratio of county crop shares for these crops versus the corresponding out-of-sample predicted values, all based on the regression specification in column (1) of Table 4. That is, for example, the scatter diagram labeled "Corn vs. Soybeans" plots actual $\ln(s_{corn}/s_{soybeans})$ versus predicted $\ln(s_{corn}/s_{soybeans})$ by county. The solid lines in these diagrams represent least-squares fitted values for the data in the scatter plot, while the dashed lines show a hypothetical perfect fit line for comparison.

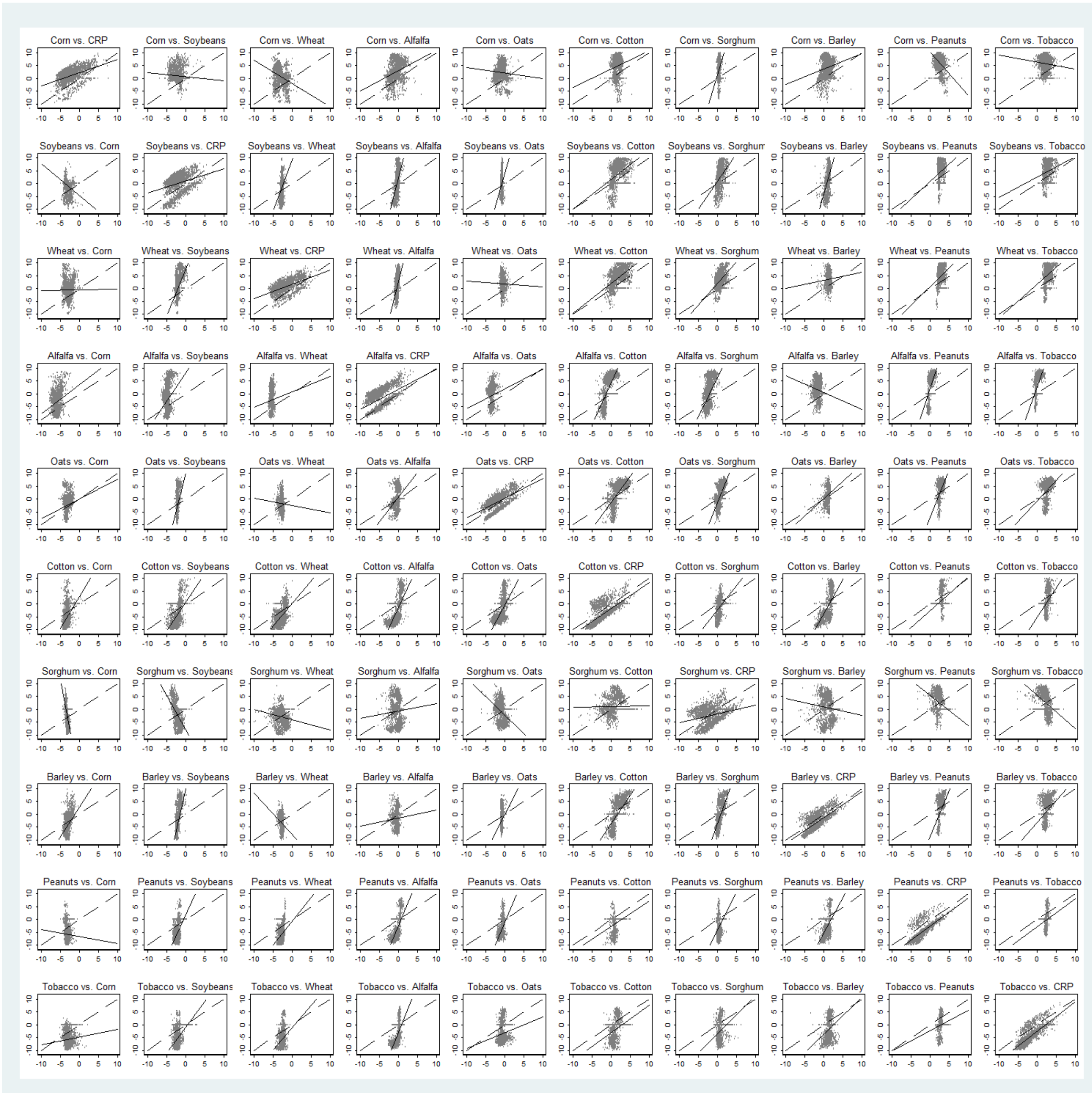

Figure 1: Out-of-sample predictions for relative crop shares

Note: Figure plots actual logged crop share ratios versus predicted logged ratios based on regressions of specification (1) that drop the crop listed first prior to estimation. Fitted values are given by the solid lines, while hypothetical perfect fit lines are given by the dashed lines. See text for details.

Importantly, in constructing our out-of-sample predicted values, we drop the crop listed first from the dataset prior to estimation and prediction. That is, for example, the scatter diagram labeled "Corn vs. Soybeans" is based on a regression that excludes corn from the estimation sample. The array presents one such diagram for every pairing of the ten most popular crops, as well as for these crops paired with CRP land (on the diagonal).

In all but 18 of the 100 unique pairings presented in this figure, the correlation between actual and predicted relative crop shares is positive, which indicates that our model is doing a fairly good job of explaining broad patterns in relative crop shares across locations.[33] The magnitudes of these correlations are also fairly encouraging, with slopes near one in many cases, implying relatively unbiased out-of-sample forecasts.[34] If we focus instead on overall forecast accuracy, we see that the variation in predicted relative crop shares (the horizontal spread) is substantially less than the variation in actual relative crop shares (the vertical spread). This is consistent with the R-squared measure in column (1) of Table 4, which indicates that the crop-county interactions explain just 8% of the variation in relative crop shares across counties (after removing crop and county dummies). Thus, while actual and predicted crop shares correlate as expected in most cases, actual shares could deviate from predicted shares quite substantially in particular counties. When we replicate this analysis for regression specification (3), which adds crop-region fixed effects, and specification (4), which limits the choice set to the ten most popular crops, our conclusions are essentially the same as before. While actual and predicted shares remain broadly consistent on average, the deviation in any particular county may be quite substantial.

6 Simulating crop displacement

In this section, we simulate the effects of increased demand for cellulosic ethanol to meet the federal RFS on the displacement of conventional food crops. We use our conceptual model in equation (2), the empirical model in equation (4), and our econometric estimates of this model in column (1) of Table 4 to predict county and aggregate crop shares when switchgrass

[33] Most of these exceptions involve either corn or sorghum.

[34] For logged crop ratios involving CRP land along the diagonal, the near-perfect correlations largely reflect county dummies, which capture each county's overall suitability for growing crops relative to CRP. For the remaining pairings, however, the county fixed effects cancel out. Thus, these other correlations are driven entirely by the interactions between county characteristics and crop attributes.

and miscanthus are added to the choice set and are sufficiently profitable to be adopted by many landowners. As shown in our conceptual model, rising output prices are captured fully by the crop dummies. Thus, we simulate the effects of increased feedstock demand by gradually manipulating the coefficient on the switchgrass or miscanthus dummy, holding all else equal, until aggregate production is just sufficient to meet the RFS, calculating the displacement of other crops in response.

Our simulations are based on increasing the return for domestic production of switchgrass or miscanthus, while implicitly holding returns for other crops fixed at their 1986–2008 historical levels. In the case of sugar cane, sugar beets, and peanuts, for which realized returns likely exceed marginal opportunity costs during this time period due to restrictions on production quantities, we implicitly assume that returns are held fixed at marginal opportunity costs, with farmers freely able to move land in and out of different crops without constraint. In addition, we implicitly assume zero ethanol imports to meet the RFS requirements. This partial-equilibrium analysis obviously ignores the potential effects of rising feedstock production on food prices, rising domestic ethanol prices on trade, and interactions with commodity price support programs. A more complete analysis would require a model of crop demand, energy markets, and agricultural policy, however, which is beyond the scope of this paper. Also note that since our model restricts the choice set to cropland, we ignore potential impacts on other land uses, including forest and pasture.

The presence of the crop-county residuals (the ξs) in our choice model presents a dilemma for simulating the adoption of new crops, since we have no direct estimate of these residuals for new crops. Thus, as one benchmark, we simulate crop displacement and feedstock supply assuming that the residual terms for new crops are zero in all counties. As a second benchmark, we use alfalfa's residuals to proxy for those of switchgrass and miscanthus. This is a sensible alternative approach, since alfalfa has much in common with these dedicated bioenergy feedstocks. It is a perennial crop that once established can be harvested for multiple years without replanting, it is often grown on less productive land, and during harvest it is cut, left in the field to dry, baled, and then transported relatively short distances for local consumption. These features align closely with feedstock production and transportation to nearby bioenergy processing facilities.[35]

Table 5 calculates the displacement of food crops when bioenergy feedstocks comprise 6.6% of all cropland, which is roughly consistent with using switchgrass to meet the federal RFS of 16 billion gallons of cellulosic ethanol in 2022.[36] In absolute terms, crop shares tend to decrease most for the most widely adopted crops, including corn, soybeans, wheat, CRP land, and alfalfa, regardless of what we assume about the residual terms for new crops. In relative terms, however, bioenergy feedstocks tend to displace crops with similar attributes, which compete for land with similar soil and climate characteristics. Among the ten most popular crops, switchgrass and miscanthus have disproportionate impacts on sunflowers, barley, wheat, alfalfa, oats, and CRP land, with somewhat milder impacts on corn, soybeans, cotton, and sorghum. When we assume the same county residuals as alfalfa, these impacts shift toward alfalfa and other crops that are grown in the same counties as alfalfa, but overall, the results are qualitatively similar.

Figure 2 maps our model's predictions for where switchgrass and miscanthus are likely to be grown. While both crops are widely grown, their production tends to concentrate in the northern plains. This is consistent with our crop-by-crop displacement results above, which show larger impacts on sunflowers, barley, wheat, and other crops grown in the northern plains. Little switchgrass or miscanthus is grown in the southwest or northeast since these are agriculturally unproductive areas, with poor soil and climate conditions for crop production

[35]One potential weakness of this approach is that alfalfa is valued partly for its ability to fix nitrogen, while switchgrass and miscanthus are not. We have considered adding hay to our model and using hay's residuals instead, but hay is an amalgam of different grass species that vary over time and across locations. Another potential weakness is that data on alfalfa acres are not available for most Southern and a few Midwestern states. Thus, alfalfa's residuals in these counties are determined largely by our ad-hoc fix of the zero shares problem.

[36]Assuming a switchgrasss yield of 9 tons per acre, a conversion rate of 90 gallons per ton, and 300 million acres of total cropland, producing 16 billion gallons of cellulosic ethanol domestically would require roughly 20 million acres or 6.6% of all cropland. Per-acre yields for miscanthus are projected to be considerably higher than 9 tons per acre. We assume the same per-acre yield as switchgrass to facilitate comparison of crop displacement impacts across scenarios.

Table 5: Absolute and percent changes in crop shares

Crop	Base share	Switchgrass to meet RFS: Zero resid. Change Abs.	Pct.	Switchgrass to meet RFS: Alfalfa resid. Change Abs.	Pct.	Miscanthus to meet RFS: Zero resid. Change Abs.	Pct.	Miscanthus to meet RFS: Alfalfa resid. Change Abs.	Pct.
Switchgrass	0.0	6.6		6.6					
Miscanthus						6.6		6.6	
Alfalfa	5.6	-0.5	-8.7	-1.0	-17.5	-0.5	-8.8	-1.0	-18.0
Barley	2.2	-0.3	-12.7	-0.2	-7.5	-0.3	-11.4	-0.2	-7.5
Beans	0.5	-0.1	-12.9	-0.0	-4.4	-0.1	-11.2	-0.0	-4.1
CRP	9.5	-0.7	-7.9	-0.6	-6.7	-0.7	-7.6	-0.6	-6.6
Chickpeas	0.0	-0.0	-9.1	-0.0	-3.8	-0.0	-11.0	-0.0	-4.0
Corn	26.4	-1.1	-4.4	-1.8	-6.7	-1.2	-4.5	-1.7	-6.6
Cotton	3.9	-0.2	-4.7	-0.0	-1.0	-0.3	-6.6	-0.0	-1.2
Flaxseed	0.2	-0.0	-19.4	-0.0	-6.3	-0.0	-16.5	-0.0	-5.5
Mustard	0.0	-0.0	-14.0	-0.0	-5.9	-0.0	-13.2	-0.0	-5.9
Oats	2.5	-0.2	-7.9	-0.2	-9.5	-0.2	-8.0	-0.2	-9.2
Peanuts	0.5	-0.0	-9.7	-0.0	-1.7	-0.0	-10.9	-0.0	-1.9
Peas	0.2	-0.0	-18.8	-0.0	-5.0	-0.0	-16.6	-0.0	-4.5
Potatoes	0.3	-0.0	-9.3	-0.0	-7.5	-0.0	-10.0	-0.0	-9.2
Rapeseed	0.4	-0.1	-20.8	-0.0	-4.1	-0.1	-16.8	-0.0	-3.3
Rye	0.3	-0.0	-13.0	-0.0	-8.5	-0.0	-13.4	-0.0	-8.8
Safflower	0.0	-0.0	-15.1	-0.0	-6.7	-0.0	-14.4	-0.0	-6.4
Sorghum	3.3	-0.2	-6.1	-0.2	-6.6	-0.2	-6.9	-0.2	-6.6
Soybeans	21.6	-0.9	-4.1	-0.9	-4.0	-0.9	-4.2	-0.9	-4.0
Sugarbeet	0.5	-0.1	-11.7	-0.0	-4.7	-0.0	-10.3	-0.0	-4.9
Sugarcane	0.1	-0.0	-5.7	-0.0	-1.0	-0.0	-7.6	-0.0	-1.1
Sunflower	0.9	-0.1	-15.5	-0.1	-7.5	-0.1	-13.6	-0.1	-7.0
Tobacco	0.2	-0.0	-9.9	-0.0	-5.6	-0.0	-11.8	-0.0	-6.6
Tomatoes	0.1	-0.0	-4.2	-0.0	-0.6	-0.0	-7.4	-0.0	-0.6
Wheat	20.9	-1.9	-9.1	-1.5	-7.0	-1.8	-8.7	-1.5	-7.1
Food & CRP	100.0	-6.6	-6.6	-6.6	-6.6	-6.6	-6.6	-6.6	-6.6
Food	90.5	-5.8	-6.5	-5.9	-6.6	-5.9	-6.5	-6.0	-6.6
Top-4 food	74.5	-4.4	-5.9	-5.1	-6.8	-4.4	-5.9	-5.1	-6.8

Note: Table reports displacement of current crops when either switchgrass or miscanthus are added to the choice set at a level of profitability such that they comprise exactly 6.6% of cropland, which is roughly consistent with the number of switchgrass acres required to meet the federal RFS. The first column of results reports baseline crop shares. The remaining columns report both absolute and percent changes in crop shares when either switchgrass or miscanthus are used to meet the federal RFS and when we either set the county residuals for these feedstocks to equal zero or to equal alfalfa's residuals. Rows at bottom calculate shares across different categories of cropland. See text for details.

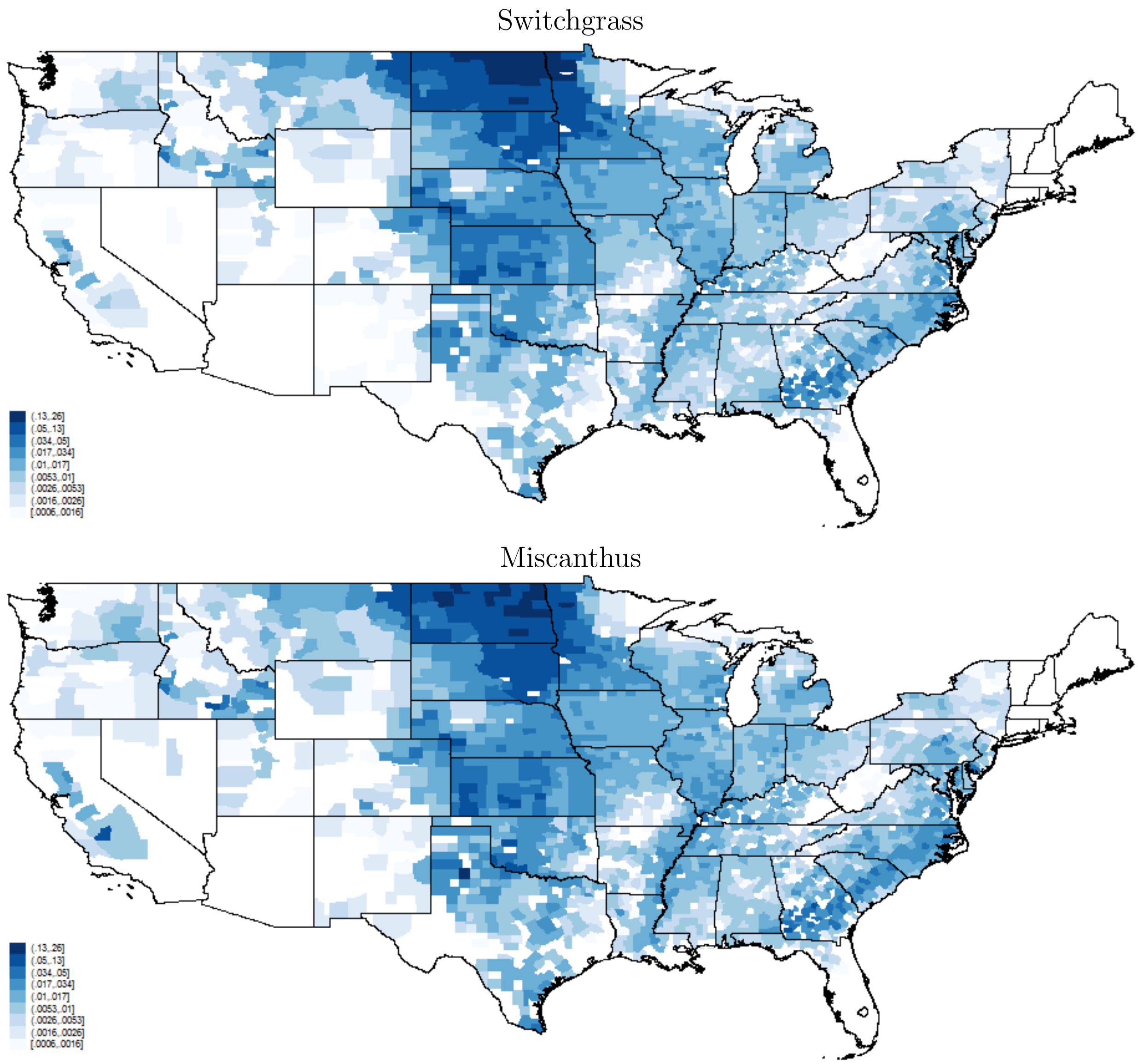

Figure 2: Forecast land-use shares for bioenergy crops

Note: Figures map county land-use shares for switchgrass and miscanthus when these feedstocks comprise 6.6% of all cropland. Darker blue shading corresponds to higher feedstock shares, as indicated in the legend; white indicates missing data (typically non-agricultural counties). See text for details.

in general.

Our results imply that meeting the federal RFS exclusively through the domestic production of miscanthus would likely have a smaller impact on world food prices than would using switchgrass. While we forecast that both crops would have roughly the same proportional impact on major food crops, miscanthus yields are projected to be nearly twice as high as those of switchgrass. Thus, meeting the federal RFS exclusively through the use of switchgrass would require much more cropland.

Our results also highlight an important point about crop choice modeling. Obviously, crops with similar attributes tend to compete for the same land. We see this manifest in our model as similar patterns of adoption for switchgrass and miscanthus, whose attributes happen to be similar. Less obvious is the crucial role played by the joint distribution of soil and climate characteristics and the unintuitive patterns of crop choice that can sometimes result. For example, switchgrass and miscanthus are more cold tolerant and more sensitive to rainfall than most other crops, suggesting that they should outperform other crops in cold, rainy climates, all else equal. Yet the coldest agricultural areas in the United States also tend to be dry—at least in the eastern half of the country. Faced with this tradeoff between rain and cold, our model predicts that switchgrass and miscanthus will opt for the cold, displacing wheat and other crops grown in the northern plains. In other regions, however, with different correlations between soil and climate characteristics, patterns of crop displacement could be very different.

7 Conclusion

We develop an econometric model of crop choice in which a farmer's profit-maximizing crop depends on crop attributes interacted with local soil and climate characteristics. We estimate the model using county-level data on average crop choices over the last two decades. Attributes of chosen crops correlate as expected with soil and climate characteristics. For example, crops that use water efficiently tend to be grown in dry regions, while crops with long growing seasons tend to be grown in warmer climates. This model allows us to forecast the adoption of new bioenergy crops in response to cellulosic ethanol mandates and rising bioenergy feedstock prices. Our results indicate that switchgrass and miscanthus would both tend to be adopted in the northern plains, displacing wheat and other crops grown in that region, with somewhat milder impacts on corn and soybeans. Yet the use of switchgrass would likely have a disproportionate impact on world food prices, given its comparatively low yields. Our results could be used to validate the findings of simulation models or to help parameterize a model of U.S. crop supply, while our overall approach could easily be extended to include additional crops and countries.

Our model and empirical estimates have several limitations. First, we only model crop choice conditional on growing crops. We do not model the choice of cropland versus other land uses, such as forest, pasture, and rangeland. In future work, we plan to extend our model to include these other land uses, as in Lubowski, Plantinga, and Stavins (2008). Second, we impose a fair bit of structure on both our conceptual and empirical models, including the assumption that variable inputs substitute easily for fixed inputs, as well as the assumption of iid errors in our main specification. In future work, we plan to relax these restrictions. Third, we have not yet been able to recover the full crop supply function from our estimates due to a small number of crops and an even smaller number of crops for which the USDA tracks financial returns. In principle, however, we should be able to recover the supply function from our estimated crop dummies, perhaps using time-series variation in crop prices, and we are actively pursuing such approaches. Fourth, while we have treated the accumulated knowledge about crop attributes as data, we believe there is considerable room for improvement in quantifying crop attributes rigorously and consistently across crops using modern empirical techniques, which we plan to pursue in future research. Finally, our out-of-sample validation exercise indicates that while observed crop attributes provide a good deal of information about the adoption of new crops, there is much room for improvement. Thus, we remain humble about our ability to forecast the adoption of new crops with precision.

In spite of these glass-half-empty, glass-half-full results, we emphasize three attractive features of our econometric approach relative to existing simulation-based approaches in the literature. First, since we estimate the parameters of our model directly from observed crop choices, we argue that our model has the potential to predict the adoption of new crops more accurately by better capturing the actual behavior of farmers attempting to maximize profits when faced with different growing conditions. Second, rather than sweep our model's prediction errors under the rug, we have laid them bare. Modelers that impose constraints on county crop shares to ensure that predicted values more closely match observed values may only be obscuring deficiencies in their models. We interpret our prediction errors economically as reflecting unobserved costs and benefits that shift relative crop returns. In contrast, imposing constraints inappropriately implies that crop choice is perfectly inelastic to price increases at the constraint. Third, while our simple model undoubtedly misses many important aspects of crop production, and while we are actively working to refine and extend our model, the parameters we estimate, their interpretation, and the variation in the data we use to identify them are all abundantly clear. This is not true of many bottom-up simulation models, which often contain hundreds of parameters, and for which the methodology used to choose the parameters is not always clear.

References

Albouy, D., W. Graf, R. Kellogg, and H. Wolff (2011). Climate amenities, climate change, and American quality of life. Working Paper.

Asrat, S., M. Yesuf, F. Carlsson, and E. Wale (2010). Farmers' preferences for crop variety traits: Lessons for on-farm conservation and technology adoption. *Ecological Economics 69*(12), 2394 – 2401.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica 63*(4), pp. 841–890.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics 25*(2), 242–262.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *The Review of Economic Studies 47*(1), 225–238.

Chen, X., H. Huang, M. Khanna, and H. Önal (2011, January). Meeting the mandate for biofuels: Implications for land use, food and fuel prices. NBER Working Paper 16697.

Dalton, T. J. (2004). A household hedonic model of rice traits: Economic values from farmers in West Africa. *Agricultural Economics 31*(2-3), 149–159.

Deschnes, O. and M. Greenstone (2007). The economic impacts of climate change: Evidence from agricultural output and random fluctuations in weather. *The American Economic Review 97*(1), pp. 354–385.

Egbendewe-Mondzozo, A., S. Swinton, R. Izaurralde, D. Manowitz, and X. Zhang (2011). Biomass supply from alternative cellulosic crops and crop residues: A spatially explicit bioeconomic modeling approach. *Biomass and Bioenergy*.

Ekanem, E. P. and W. B. Sundquist (1993). Estimating marginal implicit prices for selected quality attributes of hybrid seed corn. Staff Papers 13473, University of Minnesota, Department of Applied Economics.

Ferris, J. and J. Siikamki (2009, August). Conservation Reserve Program and Wetland Reserve Program. Backgrounder, Resources for the Future.

Fisher, A. C., W. M. Hanemann, M. J. Roberts, and W. Schlenker (2011). The economic impacts of climate change: Evidence from agricultural output and random fluctuations in weather: Comment. *American Economic Review Forthcoming*.

Holland, S. P., J. E. Hughes, C. R. Knittel, and N. C. Parker (2011, September). Some inconvenient truths about climate change policy: The distributional impacts of transportation policies. NBER Working Paper 17386.

Khanna, M., X. Chen, H. Huang, and H. Onal (2011). Supply of cellulosic biofuel feedstocks and regional production pattern. *American Journal of Agricultural Economics 93*(2), 473–480.

Kiniry, J., C. Jones, J. O'toole, R. Blanchet, M. Cabelguenne, and D. Spanel (1989). Radiation-use efficiency in biomass accumulation prior to grain-filling for five grain-crop species. *Field Crops Research 20*(1), 51–64.

Lubowski, R. N., A. J. Plantinga, and R. N. Stavins (2008). What drives land-use changes in the United States? A national analysis of landowner decisions. *Land Economics 84*(4), 529–550.

Massetti, E. and R. Mendelsohn (2011, June). Estimating ricardian models with panel data. NBER Working Paper 17101.

Mendelsohn, R., W. D. Nordhaus, and D. Shaw (1994). The impact of global warming on agriculture: A ricardian analysis. *The American Economic Review 84*(4), 753–771.

Mullahy, J. (2010, September). Multivariate fractional regression estimation of econometric share models. NBER Working Paper 16354.

Mundlak, Y. (1978, January). On the pooling of time series and cross section data. *Econometrica 46*(1), 69–85.

Nevo, A. (2000). Mergers with differentiated products: The case of the ready-to-eat cereal industry. *The RAND Journal of Economics 31*(3), pp. 395–421.

Papke, L. E. and J. M. Wooldridge (1996, November-December). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics 11*(6), 619–32.

Petrin, A. (2002). Quantifying the benefits of new products: The case of the minivan. *Journal of Political Economy 110*(4), pp. 705–729.

Roberts, M. J. and W. Schlenker (2010a, April). Identifying supply and demand elasticities of agricultural commodities: Implications for the us ethanol mandate. NBER Working Paper 15921.

Roberts, M. J. and W. Schlenker (2010b, August). Is agricultural production becoming more or less sensitive to extreme heat? Evidence from u.s. corn and soybean yields. NBER Working Paper 16308.

Schlenker, W., W. M. Hanemann, and A. C. Fisher (2005). Will u.s. agriculture really benefit from global warming? accounting for irrigation in the hedonic approach. *The American Economic Review 95*(1), pp. 395–406.

Schlenker, W., W. M. Hanemann, and A. C. Fisher (2006, May). The impact of global warming on U.S. agriculture: An econometric analysis of optimal growing conditions. *The Review of Economics and Statistics 88*(1), 113–125.

Schlenker, W. and M. J. Roberts (2009). Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proceedings of the National Academy of Sciences*.

Seo, S. N. and R. Mendelsohn (2008). An analysis of crop choice: Adapting to climate change in South American farms. *Ecological Economics 67*(1), 109–116.

Timmins, C. (2002, September). Measuring the value of varietal technology in Brazilian agriculture. Unpublished paper.

Timmins, C. (2006). Endogenous land use and the ricardian valuation of climate change. *Environmental and Resource Economics 33*, 119–142.

Train, K. E. (2009). *Discrete Choice Methods with Simulation* (Second ed.). Cambridge University Press.