Using continental grids to improve our understanding of global land supply responses: Implications for policy-driven land use changes in the Americas

By

Nelson Villoria and Jing Liu*

GTAP Working Paper No. 81

June 2015

* Villoria (corresponding author) is a Research Assistant Professor and Liu is a Post-Doctoral Fellow, both with the Department of Agricultural Economics, Purdue University. Please address correspondence to Nelson Villoria: nvilloria@gmail.com. This work has been possible due to generous support from the Energy Policy Research Institute, the Geospatial Building Blocks project (NSF Grant #1261727), and the GEOSHARE project.

Using continental grids to improve our understanding of global land supply responses: Implications for policy-driven land use changes in the Americas

Nelson Villoria and Jing Liu*

June 25, 2015

Abstract

Global economic models with explicit treatment of global land markets are crucial to understanding the consequences of different policy choices on global food and environmental security. However, these models rely on parameters for which there is little econometric evidence. A fundamental parameter in these models is the land supply elasticity. We provide a novel set of land supply elasticities estimated using gridded data for the American continent, and we use them in exploring previous work on the indirect land-use effects of US ethanol policy. Our estimates provide a basis for better-informed simulations of global land-use transitions under different economic and policy scenarios.

JEL Codes: Q24, C21, C68

*Villoria (corresponding author) is a Research Assistant Professor and Liu is a Post-Doctoral Fellow, both with the Department of Agricultural Economics, Purdue University. Please address correspondence to Nelson Villoria: *nvilloria@gmail.com*. This work has been possible due to generous support from the Energy Policy Research Institute, the Geospatial Building Blocks project (NSF Grant #1261727), and the GEOSHARE project.

The conversion of forests to cropping and grazing lands accounts for under a quarter of global anthropogenic greenhouse gas emissions (Smith et al., 2014, p. 816). At the same time, forests and other land uses offset up to 20% of those emmissions by absorbing atmospheric carbon during photosynthesis (Tubiello et al., 2014, p. 19). The dual role of land-based activities as emitters and sinks of GHGE has led policy makers and researchers to carefully scrutinize policies that may alter global patterns of land use.

Quantitative models of global commodity markets are an essential part of this scrutiny. These models permit isolating the effects of a specific policy from the myriad other factors that affect the economy over a given period of time (Babcock, 2009). The effects of these policies on prices and quantities can then be systematically tracked as they move throug the supply systems of many regions connected via international trade. Once the supply effects of the policy under study are resolved, researchers combine changes in land use with GHGE factors to calculate the net changes in emissions (Hertel et al., 2010; Dumortier et al., 2011).

The strategy just described has been particularly prominent in the U.S. where model predictions of indirect land use effects and associated GHGE emissions are used to regulate the ethanol industry (Babcock, 2009). Beyond biofuels, global trade models have been used to model the land use changes and associated GHGE of natural covers (Hertel et al., 2009), technological change (see Villoria, Byerlee, and Stevenson, 2014, for a recent review), international trade (van Meijl et al., 2006; Verburg et al., 2009), climate change mitigation (Golub et al., 2009), and agricultural policies (Eickhout et al., 2007).

Explicitly or implicitly, these models rely on assumptions about the ease with which non-cropland covers—such as savannas and forests—convert to cropland in response to changes in the relative profitability of alternative uses. Two contrasting approaches are offered by the FAPRI/CARD and GTAP-AEZ models, arguably the most widely used models to determine the indirect land use effects of biofuels in the U.S. (Babcock, 2009; Gohin, 2014) and described in Hayes et al. (2009) and Hertel et al. (2009). The FAPRI/CARD model assumes that cropland expansion into forests and grasslands occurs at historically observed rates (Searchinger et al., 2008; Dumortier et al., 2011). In contrast, the GTAP-AEZ framework assumes a fixed land endowment that is transformed across natural covers as dictated by a constant elasticity of transformation (Hertel et al., 2009).

Other approaches to modeling land use do exist, for example, in Gurgel, Reilly, and Paltsev (2007), Eickhout et al. (2007), or Jones and Sands (2013); they all make assumptions about the responsiveness of natural lands to changes in agricultural returns. Unfortunately, and in contrast to total area (e.g., Roberts and Schlenker, 2013) or cropland specific elasticities (see Miao, Khanna, and Huang, 2015, for a recent review and new estimates), estimates of land transition responses are scarce (Barr et al., 2011; Taheripour and Tyner, 2013), and to the best of our knowledge, concentrated on the U.S. (Lubowski, 2002; Rashford, Walker, and Bastian, 2011; Scott, 2013).

The concentration of land conversion estimates in the U.S. weakens the empirical basis underpinning available global analysis, compounding the uncertainty that arises from other crucial parameters such as the response of yield to prices (Gohin, 2014), crop-specific area elasticities (Keeney and Hertel, 2009), international trade responses (Villoria and Hertel, 2011), productivity of marginal lands (Stevenson et al., 2013), and possibilities for land intensification such as multiple cropping of the same plot during the growing season (Babcock and Iqbal, 2014). Indeed, in a Monte Carlo study, Lobell, Baldos, and Hertel (2013) found that the uncertainty in land supply elasticities dominates the uncertainty from all of the aforementioned sources, except the possibility for land intensification, which is not included in their model.

Against this background, the objective of this paper is to estimate a set of land conversion elasticities that takes into account the heterogeneity in land supply responses across countries. Our focus is on the contiguous Americas, from Canada to Argentina. Differently from Lubowski (2002), Rashford, Walker, and Bastian (2011), and Scott (2013), this large-scale coverage forces us to work in a data sparse environment. We work around these data limitations by using a spatially explicit model of land use based on Chomitz and Gray (1996), estimated using continental gridded datasets on a limited set of biophysical, agronomic, and socioeconomic variables. Our dependent variable is a binary choice between non-cropland and cropland. At the continental level, we do not observe land returns. But we observe the access to markets of each parcel (Verburg, Ellis, and Letourneau, 2011). The theoretical model maps the changes in accessibility of any parcel onto changes in parcel-level net agricultural returns. This implication is a testable hypothesis, and, along with the property of the logit model to produce marginal effects that are observation specific, allows us to calculate land conversion elasticities—defined as the relative change in the probability of transitioning from non-cropland to cropland given a relative change in agricultural net returns—for each gridcell in the sample.

To preview the results, table 1 compares our estimates to those of Lubowski (2002). We use the underlying potential vegetation of each grid-cell from Ramankutty and Foley (1999) to identify whether the transition is from forests or grasslands. We weigh each grid-cell's land conversion elasticity by its fitted probability and sum up a region roughly comparable to Lubowski (2002)'s. While we postpone the discussion about the details of this aggregation to the results section, note that for forests, our estimates are very close (around 0.3), although our pasture to cropland elasticity is considerably lower (ours: 0.23, Lubowski, 2002, : 0.34). We then use our estimates to calibrate the own-price cropland elasticity and CET parameters for the U.S. following Ahmed, Hertel, and Lubowski (2008), which are the analogs of the current parameters in the GTAP model: Table 1 shows that these are virtually indistinguishable from each other.

Table 1: Existing estimates for the U.S. compared to our results.

Author	Crop-return elasticity	From ... to cropland	Value	Our estimate
Lubowski	of transition probabilities	US managed pastures	0.34	0.23
		US managed forests	0.30	0.33
Schott	of non-cropland to cropland	US managed non-croplands	0.30	0.26[1]
Ahmed et al.	of cropland & CET parameter	cropland	0.06 -0.20	0.06 -0.22

[1] Percentage change in land supplied to agriculture when we increase market access by 1%.

In order to get proper land supply elasticities, we elicit a land supply schedule by using the gridcell level conversion elasticities to update the fitted probabilities of each gridcell as net returns increase. The updated probabilities determine whether the parcel transitions from non-cropland to cropland. By adding all the parcels that make this transition at each level of net returns, we obtain a land supply schedule. We use these schedules to calculate region-specific land supply arc-elasticities for different regions in the continent. The US arc-elasticity, corresponding to a 10% increase in agricultural returns, is displayed next to Scott (2013)'s result. As in the case of the land conversion elasticities, our estimate is virtually indistinguishable from his, which was obtained using a much richer dataset.

Encouraged by these results, we revisit the study of indirect land use effects of US biofuels by Hertel et al. (2010). This study underpins the land use thresholds used by the California Air Resources Board in order to qualify a biofuel source as sustainable, and therefore is of policy relevance (Babcock and Iqbal, 2014). When we recalibrate their model to our new estimates, we find that the non-US land supply response is considerably overestimated. This finding is in line with the recent discussion by Babcock and Iqbal (2014), who found that both FAPRI and GTAP likely overestimate land use changes by not considering land intensification possibilities (e.g., double cropping) which are more likely to occur due to the high costs of converting new land into agriculture. Our findings, therefore, underscore the importance of taking into account regional differences in global land supply responses. As our work demonstrates, this can be achieved in a relatively data sparse environment that exploits the spatial nature of large-scale grids using the guidance of a simple economic model. Along with these insights, the disaggregated elasticities as well as the data to calculate them at varying levels of aggregation comprise the main contributions of this article to the existing literature.

A Deeper Look at Existing Elasticities

Lubowski (2002) estimates land-use transition probabilities among a set of six possible uses: crop, pasture, forest, range, urban, and the Conservation Reserve Program. For this, he combines spatially-explicit, land-plot level data on land use from the US National Resource Inventory with county-level net returns based on commodity prices, government payments, and other sources of land-use revenues and production costs. His main contribution is estimates of the elasticities of land-use transition probabilities with respect to changes in the relative returns of the alternative uses. Table 1 displays two of these estimates, namely the relative changes in the probability of transitioning from either pastures or forests to cropland as a response to changes in agricultural net returns, keeping the returns to the alternative use constant. These two values from Lubowski (2002) are used by Ahmed, Hertel, and Lubowski (2008) to calibrate the own-price land supply elasticity (0.06 in table 1) that underpins the elasticity of transformation (-0.2) assumed to be constant across all the regions in most of the published work that uses the GTAP-AEZ framework (e.g., Hertel et al., 2010).

Rashford, Walker, and Bastian (2011) use similar data, but they focus on grassland conversion in the Prairie Pothole region. Similar to Lubowski (2002), Scott (2013) focuses on the U.S. as a whole.

He uses a dynamic choice model that also relies on very detailed spatially explicit land uses over time (2006-2012) from the National Agricultural Statistics Service's Cropland Data Layer. Similar to Lubowski (2002), Scott (2013) also constructs county-level measures of agricultural net returns using data on prices, production costs, and subsidies. In contrast to Lubowski (2002), he lumps together all non-cropland uses and assumes that the relevant return for them is the rental rates of pastures. Instead of conversion elasticities, he provides a long-run, total acreage elasticity of 0.3 (also displayed in table 1).

Other recent estimates of total acreage elasticities for the U.S. are provided by Roberts and Schlenker (2013), who report a range of 0.26-0.33, which is in line with Scott (2013)'s findings; these authors also report total acreage elasticities for Brazil (0.22-0.38), China (0.03-0.07), India (0.01-0.02), and Thailand (0.10-0.13). Barr et al. (2011) also calculated a total acreage elasticity for the U.S., although their value of 0.02 is much lower than other estimates, which is probably explained by their focus on changes during a very short time period (from 2004-2006 to 2007-2009). Barr et al. (2011) also reports total acreage elasticities for Brazil, ranging from 0.2 (1997-99 to 2001-03) to 0.007 using the shorter time period of 2004-06, to 0.08 during 2006-2009. The main limitation of these total acreage elasticities is that they are silent on which land cover transitions into or out of agriculture, which is a key piece of information for the estimation of GHGE. In the absence of this information, modelers are forced to make ad-hoc assumptions about which land conversion is taking place; for example Searchinger et al. (2008) assumed that 36% of total cropland expansion in the U.S. would come from forests, a hypothesis rejected by Dumortier et al. (2011), who found that in most US states, the main source of cropland expansion is idle cropland.

Modeling Framework

We start with a version of the model by Chomitz and Gray (1996) in which there are only two land uses: agriculture and natural state. Formally, let Y_i be the agricultural output per unit of area that can be obtained in location i by using input I_i. Each location is endowed with a bundle of fixed natural factors S_i (e.g., slope, rainfall) that influence the productivity of I_i. Assuming a Cobb-Douglas technology, the production per unit of area, or yield, function is given by:

$$Y_i = S_i I_i^{\beta}, 0 < \beta < 1, \tag{1}$$

where β is the elasticity of yields to input usage.

Critical in this analysis is that output and input prices, P_i and C_i, are site specific. In the presence of these prices, standard profit maximization subject to 1 produces optimal input demands:

$$I_i^* = \left[\frac{C_i}{P_i S_i \beta}\right]^{\frac{1}{\beta-1}} \tag{2}$$

Substituting 2 back into 1 obtains optimal yields Y^*:

$$Y_i^* = S_i^{\frac{1}{1-\beta}} \left[\frac{C_i}{P_i \beta}\right]^{\frac{\beta}{\beta-1}} \tag{3}$$

Thus, the potential rent associated with devoting site i to agricultural use is $R_i = P_i Y_i^* - C I_i^*$ which, by virtue of 2 and 3, can be expressed as:

$$R_i = \left(P_i S_i C_i^{-\beta}\right)^{\frac{1}{1-\beta}} (1-\beta)/\beta. \tag{4}$$

Using 2, 3, and 4, an econometrician possessing data on site-specific yields Y_i, input usage I_i, biophysical factors S_i, output prices P_i, and input prices C_i could estimate the parameters of the production system to determine how prices, technology, and natural factors determine agricultural profitability; under the assumption that land is devoted to agriculture only if this is more profitable to do so than to devote it to any other alternative use, the econometrician could then predict land use responses to changes in prices given natural factors and technology. From a global perspective, such predictive ability would greatly improve our understanding of cross-country cropland reactions to demand and supply shocks that affect the relative returns to farming. In particular, if there is an increase in prices that could incentivize conversion of natural lands into agriculture, the framework above would allow identifying the potential rents that would accrue to a given plot, and after comparing those rents to the rents for alternative use, the analyst could decide whether the land endowment of a given country could be augmented to capture expansion into lands that are not currently in agriculture.

Unfortunately, although there are globally consistent gridded data on production, harvested area, yields, input usage, and biophysical factors (table A-1), site specific price and land rent data are unobserved. In a context similar to this, Chomitz and Gray (1996) demonstrate how these data limitations can be circumvented by exploiting data on the determinants of prices as well as on agricultural productivity using a reduced form of 4. Critical to their strategy is Von Thünen's assumption that spatial differentials in farm-gate prices are related solely to differences in transport costs to major markets, which in turn are determined by physical distance.

Following a similar strategy, we posit that output and input prices are functions of market accessibility. We use a market remoteness index from Verburg, Ellis, and Letourneau (2011) that takes into account such frictions and is based on traveling time. This measure, discussed in detail below, ranges from 0 to 100, and ranks locations according to traveling time to major cities (figure 1). An index of 100 implies that the location is very close to a major market, while 0 implies that the location is practically inaccessible. Formally, the price functions are given by:

$$\begin{aligned} P_i &= \exp\left(p + \gamma A_i\right), \gamma > 0, \\ C_i &= \exp\left(c + \delta A_i\right), \delta < 0. \end{aligned} \tag{5}$$

where A_i measures the traveling time from plot i to a relevant market, or market access; p and c are intercepts; and γ and δ indicate the percentage change in the respective prices, given a unitary change in market access; i.e, these are the semi-elasticity of prices to changes in distance. We expect output prices to be increasing in market access. Meanwhile, input prices become costlier as locations become more remote.

Using 5 to eliminate the price terms in 4, taking logs, grouping similar terms, and appending an error term yields:

$$\log(R_i) = \alpha_0 + \alpha_1 A_i + \alpha_2 \log(S_i) + \varepsilon_i. \tag{6}$$

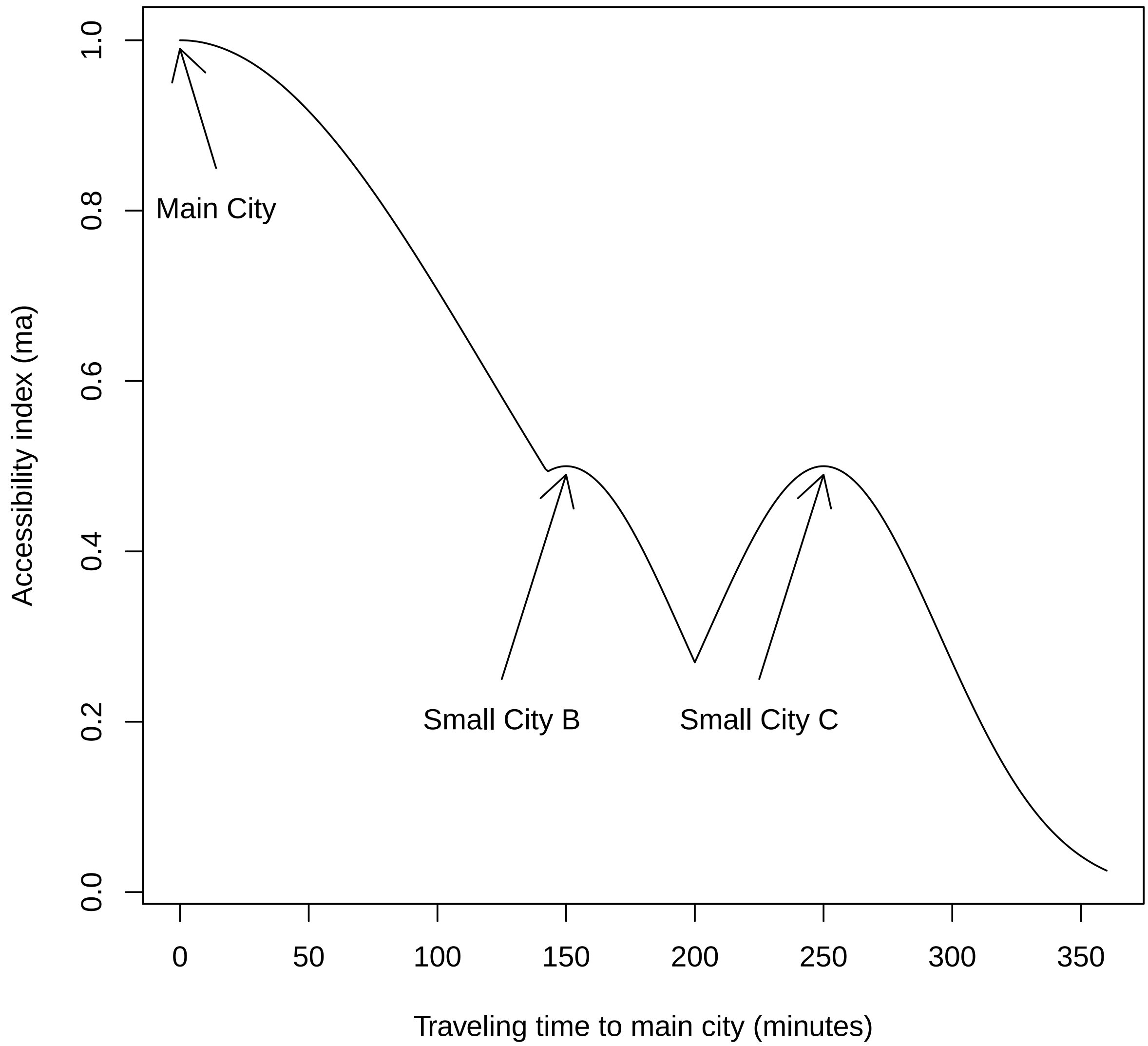

Figure 1: Market access index decreases with travel time from the location of a large market. At 150 and 250 min from the large city are two regional markets. This is figure 1 in Verburg, Ellis, and Letourneau (2011).

where $\alpha_1 = (\gamma - \beta\delta)(1-\beta)^{-1}$ is the semi-elasticity of land rents to changes in market access and $\alpha_2 = (1-\beta)^{-1}$. In principle, we would expect $\alpha_1 > 0$, a hypothesis that we test below. The last term, ε_i, is an error term whose properties are discussed in the next section.

With R_i unobservable, the following step in this strategy uses the assumption that land will be employed in the activity yielding the highest rent, a main implication of Von Thünen's model. In our case, we assume that land can either be used in agriculture or be let idle in its natural state, which produces no rent. Then, land at grid cell i will be devoted to productive use A if and only if $R_i > 0$, which is revealed in the data through the mere fact that i is used in agriculture; this allows mapping from the latent variable R_i to a binary outcome, $Z_i = 1[R_i > 0]$, which takes the value of one when production is observed, and zero otherwise. This variable can be readily constructed from global grids of agricultural production and allows estimating the parameters in model 6 using a discrete choice model:

$$P(Z_i = 1|A_i, S_i) = \Lambda(\alpha_0 + \alpha_1 A_i + \alpha_2 \log(S_i) + \varepsilon_i), \tag{7}$$

where $\Lambda(.)$ is the logistic distribution and $P(Z_i = 1|A_i, S_i)$ is the probability of observing i under agricultural production given its remoteness to markets (A_i) and natural attributes (S_i).

The partial effect of A_i on $P(Z_i = 1)$,

$$\frac{\partial P(Z_i = 1)}{\partial A_i} = \lambda\left[\alpha_0 + \alpha_1 A_i + \alpha_2 \log(S_i)\right]\alpha_1, \tag{8}$$

is the change in the probability of turning a parcel into agriculture, given a change in market remoteness. A unique aspect of this formulation is that the argument of λ, the probability distribution function of the logistic distribution, varies by plot, and thus, the partial effect of market remoteness on the probability of land use can be determined for different geographic aggregates, such as all the parcels covered by forests. The elasticity of the probability of agricultural land use to changes in market remoteness is given by:

$$\epsilon_i = \frac{\partial P(Z_i = 1)}{\partial A_i} \times \frac{A_i}{P(Z_i = 1)}. \tag{9}$$

To relate this measure back to the elasticity of land supply to change,s in input and output prices we need to map the probabilities back to a physical measure of area, say hectares, and relate changes in market remoteness to changes in input and output prices. This last linkage could be achieved by taking the ratio of input to output prices in expression 5, solving it for A_i and differencing:

$$dA_i = \frac{d\log(P_i) - d\log(C_i)}{\gamma - \delta} = dR_i. \tag{10}$$

The denominator of 10 is always positive, so the sign of dA_i depends on the signs as well as relative magnitudes of the percentage changes in output and input prices. In particular, whenever output prices grow faster than input prices, $d\log(P_i) > d\log(C_i)$, it is as if market access would increase. Likewise, if input prices grow faster than output prices it is as if market access decreases. Expression 10 highlights the fact that even though plot i and the actual market to which initial remoteness is measured against stay in the same place, market accessibility varies with price.

Moreover, equation 10 allows linking the changes in prices and costs determined in models such as GTAP to the elasticities in expression 9. A difficulty is that it is not possible to identify the semi-elasticities of input and output prices with respect to changes in market remoteness (γ and δ) from the estimates of and α_1. In the empirical application below, we set it equal to unity, although departures from this value would amplify (if less than one) or reduce (if greater than one) the effects of changes in agricultural returns on accessibility. Of more concern is the possibility of $\gamma < \delta$, which would imply that accessibility decreases as net agricultural returns increase. This could be the case, for example, if the major input is labor, and if wages in the land frontiers where conversion takes place are lower than in places closer to markets. We leave it to future research, using spatially disaggregated input and output prices, to disentangle the values of these elasticities.

By virtue of expression 10, in what follows, we refer to changes in market access as changes in land rents. We also assume that the only fixed factor of production is land; therefore, in the discussions below, we use changes in agricultural net returns and changes in land rents as synonyms.

To translate the land conversion probabilities to a proper land supply elasticity, defined as the change in the number of hectares supplied to agriculture given a change in land returns, we first consider the changes in the probability $P(Z_i = 1)$ as land rents change from an initial status $s-1$ to a subsequent status s. Formally:

$$P(Z_i = 1)_s = P(Z_i = 1)_{s-1} \times \left[1 + \epsilon_i \times \frac{\partial R_i}{R_{i,s-1}}\right] \tag{11}$$

is the updated probability of land use given a marginal change in agricultural land rents.

Following the standard practice in discrete choice modeling of using some relevant proportion in the data to determine the threshold after which a predicted or fitted probability switches from zero to one in order to predict the values of the binary dependent variable (Greene, 2008), we can further assume that parcel i shifts from non-cropland to cropland once the conversion probability surpasses a given threshold τ, which can be formalized as:

$$Z_{i,s} = \begin{cases} 1 & \text{if } P(Z_i = 1)_s > \tau \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

Each grid cell has a known physical area which we denote by L_i. Therefore, for region c (comprised by many grid cells i), the additional area brought into production as a consequence of a change in changes in returns is $L_s^c = \sum_{i \in c} Z_{is} \times L_i$. With this information, it is straightforward to construct the land supply schedule of region c by plotting L_s^c against $R_s = 1 + \partial R / R_{s-1}$ for t = 1, ... , T. The slope of the line tangent to a given point in this schedule gives us a measure of the local land supply elasticity for region r. We use these schedules to estimate the total regional area elasticity with respect to changes in land rents:

$$\eta^c = \frac{\delta L^c}{\delta R} \frac{R}{L^c}. \tag{13}$$

Notice that by using grid cell level changes in the non-cropland-to-cropland conversion probabilities, we gain a great deal of flexibility defining regions as well as the nature of the elasticities. For example, if the region is defined as all the grid cells covered by forests, $\eta_{forest,agr}$ is the cross-price elasticity of forested land supply to agriculture given a change in agricultural land returns. The usefulness of the flexibility of our estimates is demonstrated in the last section of the article where we revisit Hertel et al. (2010)'s estimates of the indirect land use effects and associate greenhouse gas emissions of US ethanol policies.

Econometric Implementation

There are three main econometric concerns for the estimation of the parameters of equation 7. First, most of the variables that we use as proxies for the natural factors S_i tend to change slowly across neighboring units, and therefore, shocks in the error term in one location are likely correlated to shocks in other locations. Moreover, many of these data are interpolations (e.g., temperature and precipitation from weather stations) and as such, subject to measurement errors likely to be systematically associated in spatial patterns, which can be captured by the model residuals (Anselin, 2002). Inferences based on parameters estimated in the presence of correlated residuals are flawed due to biased standard errors (Brady and Irwin, 2011).

Second, decisions on where to locate markets, as well as on the investments needed to reduce traveling costs from any plot to those markets may be related to the need to access the most suitable lands for agriculture (Ullman, 1941; Chomitz and Gray, 1996). This simultaneous determination of market access and land use choice could bias our parameter estimates and the land supply elasticities that we derive from them. Previous work by Chomitz and Gray (1996), looking at the role of roads in deforestation in Southern Belize, handled potential endogeneity of land use choices and road penetration by constructing a distance-to-market measure that ignored the existence of roads. This instrument also appears in the work by Nelson, Harris, and Stone (2001) which looks at roads and deforestation in the Darien, Panama. Unfortunately, this instrument does not address our concerns because it still relies on the crucial assumption that market locations are exogenous to land-use choices.

As an alternative, we alleviate endogeneity concerns by following a two-pronged strategy. First, we follow the standard practice of including in the regression a variety of land use determinants such as soil quality, climate, and topology, all dimensions of land suitability that can reasonably purge out of the error terms some of the elements of land use choices that can be correlated with market access (Chomitz and Gray, 1996; Nelson, Harris, and Stone, 2001).

In addition, we exploit the spatial nature of the data. In particular, LeSage and Pace (2009, p. 27) demonstrate that, if the omitted variables causing the correlation exhibit spatial dependence, the inclusion of spatial lagged terms for the independent variables in the so-called Spatial Durbin Model (SDM) makes the regressors orthogonal to the error term. Moreover, LeSage and Pace (2009, p. 68) empirically show that the SDM considerably reduces the bias of parameter estimates relative to ordinary least squares. The SDM is discussed in Anselin (2013), and it includes a spatial lag of the dependent variable along with the spatial lags of the independent variables. Modifying equation 6 to reflect the SDM gives rise to:

$$\log(R) = \rho W \log(R) + X\alpha + WX\alpha^* + \mu. \tag{14}$$

In this formulation, $\log(R)^*$ is the $n \times 1$ vector of log net returns to agriculture; $X = [1, A, S]$ is the $n \times k$ matrix of explanatory variables; and W is the $n \times n$ weight matrix defining neighborhood relationships among the n gridcells in the sample; $W \log(R)$ is the spatially-weighted average of latent agricultural returns in neighboring land-plots. Notice that this implies a dependency relationship whereby neighboring profitability from cropland activities influences the land-use decision. Similarly, WX are the spatially-weighted average of latent agricultural returns and explanatory factors in neighboring land-plots. In LeSage and Pace (2009)'s model, these terms capture the correlation between X and potentially omitted, spatially-dependent, omitted variables. The weight matrix is row-standardized so that letting w_{ij} denote its typical element, $\sum_j w_{ij} = 1$. This particular standardization implies that for plot i, $\sum_j w_{ij} R_j$ is the average land rent of the neighboring grid cells, weighted by the typical elements w_{ij}. The value of these weights is determined by assumptions about the strength of spatial dependence, an issue that will be revisited at estimation time. The parameter ρ is the so-called spatial lag parameter, α is a vector of parameter estimates on the effects of X, and α^* is the vector of parameter estimates on the effects of the spatially weighted X, which capture the effects of the attributes of neighboring gridcells on the land rents in plot i. The last term, μ is a $n \times 1$ vector of independently, identically, and normally distributed errors.

As mentioned above, the parameters in equation 14 can be estimated by using the observable discrete choice Z_i that results from the latent profitability. This leads us to our third concern, which is to find a suitable estimating strategy. Spatial autocorrelation is associated with heterokedastic errors, which render standard maximum likelihood estimates inconsistent (McMillen, 2006). Moreover, the likelihood function of a discrete choice variable with spatial dependence involves evaluating as many integrals as observations, which for most datasets is practically impossible due to high computation costs (Smirnov, 2010). Available Generalized Methods of Moments (GMM) estimators are robust to distributional assumptions, but they are also not feasible for large samples because each iteration requires inverting the $n \times n$ matrix $(I - \rho W)$ (Klier and McMillen, 2008).

Klier and McMillen (2008) propose an estimator that circumvents these difficulties. They derive a spatial logit GMM estimator based on the latent variable model that underlies the spatial logit. The estimator is implemented as a standard logit followed by two-stage least squares. The discrete choice model implied by their procedures is equivalent to a standard logit model estimated using a transformed X weighted by both, heterokedastic variances and the degree of spatial autocorrelation in the data:

$$P(Z_i = 1 | X_i^{**}) = \Lambda(\alpha X_i^{**} + \alpha^* W X_i^{**} + \mu_i), \tag{15}$$

where $X^{**} = X^* (I - \rho W)^{-1}$, $X_i^* = X_i \sigma_i^{-1}$, and σ_i^2 is a typical element of the diagonal of σ^2, the variance-covariance matrix of μ. The main advantage of this estimator is that it can handle large amounts of data. Monte Carlo simulations by Klier and McMillen (2008) show that the estimator accurately captures spatial effects when $\rho < 0.5$. Given these attributes, we adopt Klier and McMillen (2008)'s estimator in our work below.

Data

Table 2 reports descriptive statistics of all the variables used to estimate equation 15: table A-1 in appendix documents sources and other details of the data. The dependent variable is derived by adding up the harvested area of the 175 crops from Monfreda, Ramankutty, and Foley (2008), and assigning $Z = 1$ to any pixel which has more than 5% of its area under agriculture. This threshold is admittedly arbitrary and below we examine the robustness of our parameter estimates to different thresholds.

The market access variable comes directly from Verburg, Ellis, and Letourneau (2011), who combined spatially explicit global data on physical distance, network infrastructure, and underlying terrain to develop a high spatial resolution (1 km^2) index of market accessibility determined by traveling time from each plot to the closest and most influential market. The influence of the market is given by market size: large markets include cities with more than 750,000 inhabitants as well as maritime ports, while small markets are approximated by cities with more than 50,000 inhabitants. The authors assume that large markets are twice as important as smaller markets, and for each grid cell i in the global map, they assign a market influence index (ma_i) based on traveling time.

Figure 1 shows a large city at $T = 0$ and two small cities, B and C, 150 and 250 minutes away from the large city. The accessibility index declines as one moves farther away from the main city, and starts increasing as one gets close to city B. The plot at city B has a lower index just because it is closest to a smaller market. As one moves away from B, the accessibility index declines, reversing its trend as one gets closer to city C, just to fall again as the cities are left behind.

Potential vegetation is an important variable in our work because it identifies the natural land cover from which there is a transition to cropland. Data on grid cell level potential vegetation is from Ramankutty and Foley (1999). These data are based on satellite imagery and indicate the dominant vegetation type in each grid cell that "would likely exist now in the absence of human activities" for both cultivated and uncultivated grid cells. The original data from Ramankutty and Foley (1999) divides natural potential vegetation into 15 vegetation types, which we further aggregate to five land covers: temperate forests, tropical forests, grasslands, shrublands, and a residual category we label "other" (see table A-2 in the appendix for the correspondence between Ramankutty and Foley (1999) vegetation types).

We also add other covariates that aim to capture the role of biophysical and socioeconomic factors in the land use decision. Biophysical factors include: soil fertility constraints (6 categories: no constraints, moderate, constrained, severe, very severe, and unsuitable for cultivation) were obtained from IIASA/FAO (2012); global data on grid cell level soil organic carbon density (kg-C/m^2 to 1 m depth) and soil pH (0-14) come from the SoilData System, which was developed by the Global Soils Data Task from the International Geosphere-Biosphere Program (IGBP-DIS, 1998). These data are based on statistical resampling of global soil samples (pedon records) that are consistent with the FAO/UNESCO Soil Digital Map of the World.

Table 2: Descriptive statistics of non-cropland and cropland pixels

	Non-cropland (69% of gridcells)				Cropland (31% of gridcells)			
	Mean	s.d.	Min	Max	Mean	s.d.	Min	Max
Market access index (1-100)	7	16	0	100	23	27	0	99
Area equipped for irrigation (% of gridcell)	0	3	0	79	4	12	0	100
Precipitation (mm)	1220	862	0	7513	990	539	52	4863
Temperature (C)	17	9	-1	28	17	6	4	28
Elevation (m)	714	872	-227	5411	563	599	-25	4420
Soil fertility (IIASA classes)	5	2	1	7	3	2	1	7
Soil carbon density (kg-C/m^2)	6	3	1	24	6	2	1	22
Soil pH (0-14)	6	1	4	8	6	1	5	8
Built-up land (% of gridcell)	0	4	0	100	1	3	0	64
Protected areas	Distribution between land-use categories (%)							
Unprotected (U)			66				34	
Protected (P)			90				10	
	U			P	U			P
Distribution within land-use categories (%)	83			17	96			4
Natural potential vegetation	Distribution between land-use categories							
Shrublands (S)			74				26	
Tropical forests (Ft)			81				19	
Temperate forests (FT)			73				27	
Savannas & Grasslands (G)			49				51	
	S	Ft	FT	G	S	Ft	FT	G
Distribution within land-use category (%)	14	33	29	21	11	17	25	48

Notes: These are summary statistics for the sample of 15,093 observations (out of a total of 433,096) used to estimate the first four specifications in table 3. Details about sources and preprocessing of the data are in Appendix table A-1. Tables S-1- S-12 (supportiong online materials) report several statistics and cross-tabulations of the data that demonstrate that the sample closely mimics the information in the entire grid. Robustness tests of our main findings to different samples are discussed in the text. The omitted “Other” covers categories capturing the difference between 100 and the sum of the cover types distribution within land-use category.

The average monthly temperature (^{o}C) and average annual total precipitation (millimeters/year) over the period 1961-1990 were constructed using the data from New, Hulme, and Jones (1999). These data are commonly used for climate and ecosystem modeling and are obtained by interpolating weather station data using latitude, longitude, and elevation as predictors. The elevation data (meters above sea level) were obtained from TerrainBase, a global model of terrain and bathymetry on a regular 5-minute grid documented in NOAA (1995). We also add dummies for agroecological zones (IIASA/FAO, 2012; Monfreda, Ramankutty, and Hertel, 2009) which capture changes in the length of the growing seasons across the continent.

The socioeconomic data include area equipped for irrigation (expressed as % of the area of each 5 minute grid cell) which comes from Siebert et al. (2010). These data are based on global census-based inventory data on irrigation sources and are from the national and subnational levels. The built-up land data were obtained from SAGE and are based on observed built-up area density and nighttime lights, which in turn are used to interpolate urban-area density for those sites in which only nighttime light is observed. The 5-minute resolution data layer identifying protected areas was obtained from van Velthuizen et al. (2006).

As discussed in the next section, our choice of estimator makes it straightforward to estimate a spatial logit model using the entire continental grid: 433,096 observations below 55^{o}N after extracting islands and water bodies. However, deriving land conversion and supply elasticities requires using the parameter estimates of 15 to calculate the marginal effects and to fit the predicted probabilities. For this, we need to transform the matrix X to X^{**}, which requires inverting the matrix $(I - \rho W)$. Matrix inversion is expensive in terms of memory and forces us to draw representative samples that preserve the gridded structure. We accomplished this by sampling +/- 15,000 observations in a regular grid (3.5% of the total grid cells), which is equivalent to choosing one grid cell every 30 minutes. Regular sampling has the advantage of preserving the spatial structure of the grid, which is the preferred method when handling spatially gridded data (Cressie, 1993).

Tables S-1 S-12 in the Supporting Online Materials report several statistics and cross-tabulations of the data that demonstrate that the sample closely mimics the information in the entire grid. We also discuss the robustness of our results to the use of different samples in the next section.

Results

Table 3 reports parameter estimates for five different specifications. The first two models use Klier and McMillen (2008)'s estimator. S-I is a restricted version of equation 14 where the parameters of the spatial lags of the independent variables are assumed to equal zero. S-II displays model 14, which includes the spatial lags of both the dependent and independent variables (not displayed in the table). Column Mean S-II uses the average and standard deviations of 10,000 sample estimates of model S-II in order to test its sensitivity to the chosen sample. Columns C-I and C-II display the results of estimating S-I and S-II using a standard logit model without explicit account of spatial dependence.

Table 3: Regression results

	Spatial logit			Conventional logit	
	S-I	S-II	Mean S-II	C-I	C-II
Intercept	-5.378^*	-11.323^*	-10.755^*	-15.281^*	-12.458^*
	(2.000)	(1.765)	(1.752)	(1.543)	(1.678)
Market access	0.017^*	0.014^*	0.013^*	0.016^*	0.015^*
	(0.001)	(0.001)	(0.006)	(0.001)	(0.001)
Irrigation	0.119^*	0.128^*	0.136^*	0.122^*	0.130^*
	(0.015)	(0.016)	(0.016)	(0.008)	(0.008)
Precipitation	0.001^*	0.001^*	0.001^*	0.001^*	0.001^*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Precipitation^2	-0.000^*	-0.000^*	-0.000	-0.000^*	-0.000^*
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Temperature	0.172^*	0.567^*	0.482^*	0.755^*	0.678^*
	(0.080)	(0.062)	(0.063)	(0.041)	(0.049)
Temperature^2	-0.007^*	-0.019^*	-0.016^*	-0.025^*	-0.022^*
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Elevation	-0.000	-0.001^*	-0.001^*	-0.001^*	-0.000^*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Elevation^2	0.000	0.000^*	0.000	0.000^*	0.000^*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Soil fertility	-0.146^*	-0.150^*	-0.149^*	-0.207^*	-0.166^*
	(0.021)	(0.021)	(0.021)	(0.019)	(0.020)
Soil carbon	0.017	0.016	0.007	0.068^*	0.014
	(0.015)	(0.016)	(0.020)	(0.013)	(0.015)
Soil pH	1.015^*	1.877^*	1.826^*	2.663^*	2.193^*
	(0.548)	(0.522)	(0.508)	(0.479)	(0.491)
Soil pH^2	-0.070	-0.148^*	-0.141^*	-0.194^*	-0.176^*
	(0.043)	(0.042)	(0.041)	(0.038)	(0.039)
Built-up land	-0.051^*	-0.059^*	-0.045^*	-0.062^*	-0.061^*
	(0.007)	(0.008)	(0.007)	(0.006)	(0.007)
Protected area	-0.001^*	-0.001^*	0.000	-0.001^*	-0.001^*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ρ	0.958^*	0.377^*	0.496^*		
	(0.107)	(0.113)	(0.211)		
Pseudo-R^2	0.30	0.47		0.36	0.38

Robust standard errors in parentheses

* indicates significance at $p < 0.1$

Notes: N = 15,093. ρ in models S-I and S-II is the coefficient of spatial autocorrelation estimated on the spatial lag of the underlying, unobserved, dependent latent variable using Klier and McMillen (2008)'s estimator. S-I and S-II assume a spatial autoregressive error process. In addition, S-II includes spatial lags of all the independent variables (excluding squared terms). Spatially lagged variables include up to 6 spatial lags. Mean S-II reports the average and standard deviation of S-II regressions using 10,000 different samples. C-I and C-II are standard logit estimates of S-I and S-II.

A glance at table 3 reveals that the parameter estimates are consistent across the five specifications—both spatial and conventional—in terms of magnitude, statistical significance, and sign. The notable exception is the auto regressive parameter ρ (only estimated in S-I and S-II), which reduces almost by a factor of three when we include spatial lags of the vector of independent variables, the reason for which the model labeled S- II is our preferred specification.

To aid with the interpretation of these parameter estimates, figure 2 displays the average partial effects (APE) of each regressor on the probability of land use corresponding to the SDM S-II. These average partial effects are obtained by combining the relevant regressor (or combination of regressors) with the average scale factor evaluated at each gridcell. As explained by (Greene, 2008, p. 784), this avoids the need for estimating different marginal effects for each combination of variables handled as dummies (i.e., regions, agro-ecological zones and potential/natural vegetation), making interpretation more tractable. The upper panel of the figure shows the average partial effect for the variables with linear effects only while the lower panels display the average partial effects for the variables with a quadratic component[1].

Back on table 3, the market access index has a positive and statistically significant effect on the probability of land being used in agriculture. This is a direct test of the hypothesis of the structural model that treats market access as a proxy for land rents, and as discussed below, a pivotal finding that permits calculating land-use transition probabilities as well as land supply schedules directly linked to land returns. The average partial effect calculated using our preferred specification (figure 2) indicates that a unit increase in accessibility increases the probability of land use by around 0.17% (with a 0.14%-0.20% 95% confidence interval).

The area equipped for irrigation in each gridcell also has a positive and significant effect on the land use decision, with an average partial effect bounded by a 95% CI of 1.10%- 1.91%—such a large effect is not surprising given the enabling role of irrigation for agricultural production (Schlenker, Hanemann, and Fisher, 2005).

Precipitation and temperature show statistically significant positive linear terms and negative quadratic terms, suggesting that higher values of these variables decrease the probability of crop-land use. The marginal effects for the observed range of these two variables in the data are shown in Figure 2. For very low temperatures (in the neighborhood of zero ^{o}C), increases in heat during the growing season increase the probability of land use by around 5% (95% CI: 4.6%-8.3%). This partial effect decreases with higher temperatures, and at approximately 18^{o}C degree Celsius, higher temperatures discourage agriculture. Precipitation has much smaller effects on the probability of land use (at the mean sample value, 2000 mm/year is practically zero), although for regions with very high rainfall, the marginal effects tend to reduce the probability of agricultural land use in a more pronounced way.

[1] Algebraically, the partial effect of the i^{th} regressor on the probability of gridcell g^{th} being in agriculture is given by $\partial P(Z_g = 1|\mathbf{X_g})/\partial x_i = \lambda\left[\beta \mathbf{X_g}\right]\beta_i$, where $\mathbf{X_g}$ is gridcell g's vector of explanatory variables (including x_i) and β is the vector of regression coefficients. The average partial effect we report is given by $\sum_g \lambda\left[\beta X_g\right] G^{-1}\beta_i$, where G is the total number of gridcells. For regressors with both linear and quadratic terms, say, $\beta_{i[1]}$ an $\beta_{i[2]}$, the APE is $\sum_g \lambda\left[\beta X_g\right] G^{-1}\left(\beta_{i[1]} + 2\beta_{i[2]}x_i\right)$, which depends on a specific value of the regressor x_i. This is the reason why figure 2 displays these average partial effects for all the in-sample range of the given variable.

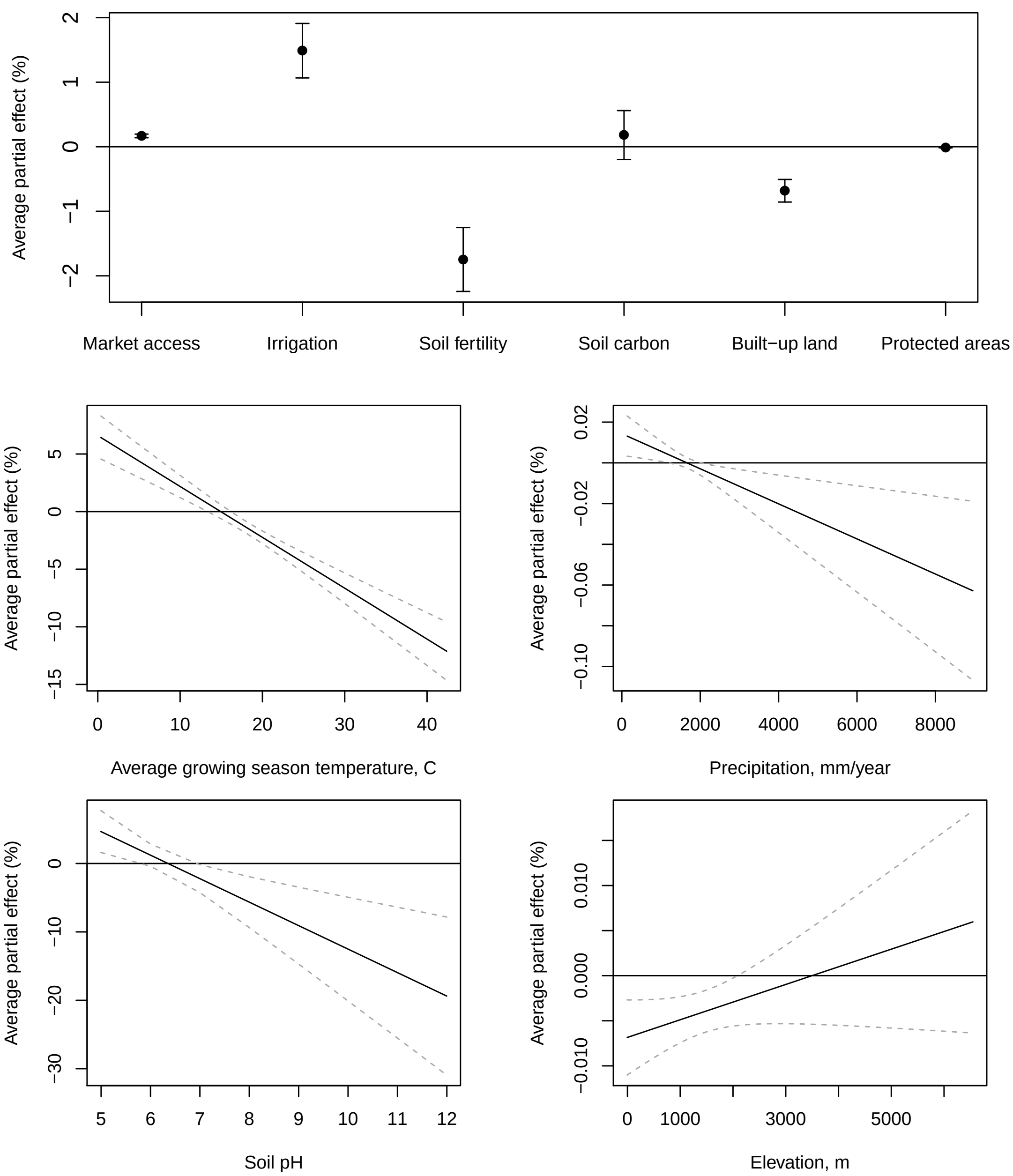

Figure 2: Average partial effects of SL-II: point estimates and 95% confidence intervals. The scale factor of the displayed partial effects are obtained by averaging the scale factors—i.e., the logit probability density functions evaluated using each gridcell's set of explanatory variables—across all the gridcells in the sample (for more details, see Greene, 2008). APE for regressors with linear effects only are shown in the upper panel, while APE for regressors with quadratic effects are shown for all the in-sample range of the variable. Standard errors were calculated using the Delta method.

The acidity of the soil (pH) also has significant linear (positive) and quadratic (negative) terms, and the APE in figure 2 suggests that the marginal effect is highest just below neutrality (ph =7) and declines as the soils become more basic. This finding is in line with the agronomic and global ecology literature, which indicates optimal growing conditions in soils with pH between 6.6 and 7.0 (Ramankutty et al., 2002).

Elevation, on the other hand, has a negative effect at first that becomes less negative as altitude increases. The average partial effects are quantitatively small. Soil fertility constraints are also significant determinants of the land use choice. More constrained soils are less likely to be used in agriculture (the APE is -1.7% with 95% CI -2.2%- -1.25%). The amount of soil carbon does not appear to significantly affect the land use decision (once we correct for spatial autocorrelation).

After correcting for accessibility, a greater share of built upland also reduces the probability of the land being used in agriculture, which probably reflects the fact that urban uses are more profitable; the effect is statistically significant with an APE of -0.68% (95% CI: -0.90%- -0.50%). We also find that although the coefficient on protected areas is statistically significant, its quantitative effects on reducing the probability of agricultural land use are small (APE = 0.01%, 95% CI: -0.02%-0.00%), suggesting that on average, law enforcement of protected parks and natural reserve are not an influential determinant of land use choices at a continental level.

Regarding goodness of fit, we use the parameter estimates of S-II to predict the probability of land use in each in-sample grid cell. The proportion of grid cells that are in agriculture is 31%. Following (Greene, 2008, p. 70-) we asume that each predicted probability that is greater than 31% is considered to be cropland and conversely, if the predicted probability is less than or equal to 31%, we consider the parcel to be non-cropland. We then compared the predicted binary variables with the observed binary variables—we found that our model correctly predicts 85% of the observations under cropland and 78% of the cases under non-cropland.

Robustness Tests

Our strategy of sampling a reduced number of observations begs the question about how representative our parameter estimates are of the underlying population parameters. In addition, estimation of the spatial logit model requires two critical choices, namely, the assumption about spatial neighborhood structures and the percentage of cropland in each gridcell required to construct the land-use choice binary variable that we aim to explain. In this subsection, we explore the robustness of our results to different empirical alternatives.

We explored the robustness of our preferred set of parameter estimates by estimating S-II using 10,000 randomly drawn samples. Table 3 (Mean S-II) shows the mean and and t-statistics of the 10,000 regressions, demonstrating that the sample used for estimation generates results both quantitatively and qualitatively comparable to those averaged over a large number of samples.

We also validate S-II by verifying its accuracy with out-of sample predictions. We performed the same procedure we used to measure the models' goodness of fit on 100 randomly drawn samples.

We found that S-II correctly predicts 82.10%-84.53% of the cases that are under agriculture (with probability thresholds ranging from 30.15% to 31.20%) and 80.54%-81.95% of the cases that are out of agriculture.

Regarding the neighborhood structure, our preferred specification is based on a neighborhood matrix that includes up to the 6 nearest neighbors. In order to test the sensitivity of our results to different choices of these parameters, the upper-part of figure 3 shows estimates of the spatial autoregressive (ρ) and market access parameters as we vary the number of neighbors from 1 to 6. As displayed by the plots, the degree of spatial autocorrelation rapidly reduces as we include more neighbors; meanwhile, the estimate of the effect of market access tends to increase as we improve our correction for spatial autocorrelation.

Regarding the sensitivity of our results to changes in the threshold of harvested area after which a gridcell is considered to be cropland (i.e., $Z_i = 1$), the lower-right plot of figure 3 shows that the proportion of gridcells which would be classified as cropland falls quickly as we increase the threshold. In other words, very few parcels would be classified as cropland if we set the threshold to, for example, 75%. Interestingly, the lower-right plot indicates that the point estimate of the effect of market access is robust to the threshold. However, as the threshold increases, the parameter estimate becomes less and less precise.

Land Supply Responses and Elasticities

In order to study the land supply response to changes in land rents, we calculated the marginal effect of changes in land rents on the land-use transition probability for each one of the 15,000 plots in our sample. We also fitted predicted probabilities of land use in each one of these 15,000 plots. Combining these two series allowed us to calculate probability-weighted average elasticities of land-use transition probabilities for different geographic aggregates of the data.

For instance, the probability-weighted average elasticities of non-crop-to-cropland transition probabilities for the entire continent is 0.13, that is, a one percent increase in net returns to agriculture causes a 0.13% increase in the probability of non-cropland shifting to cropland. However, a maintained hypothesis underlying this article is that the land supply response is heterogeneous across geographies due to differences in the accessibility and quality of the land. Therefore, it is more worthwhile to use our parameter estimates to investigate this diversity.

We start this investigation by validating our results with those in the existing literature. In the case of Lubowski (2002), we needed to subset the pixels in our sample that are within managed lands in the U.S. in order to match his spatial domain. For this, we calculated the shares of forests and grasslands that are managed using the information on the areas under pastures, cropland, and managed forest as well as areas under natural grasses and forests in Gurgel, Reilly, and Paltsev (2007, Table 3, p.13). For the U.S., Gurgel, Reilly, and Paltsev (2007) indicate that 45% of grasslands are managed pastures and 68% of forests are managed forests. We then ordered the grid cells from closest to remotest to a market. The set of ordered grid cells whose cumulative area of a given cover comprised the share of managed cover indicated by Gurgel, Reilly, and Paltsev

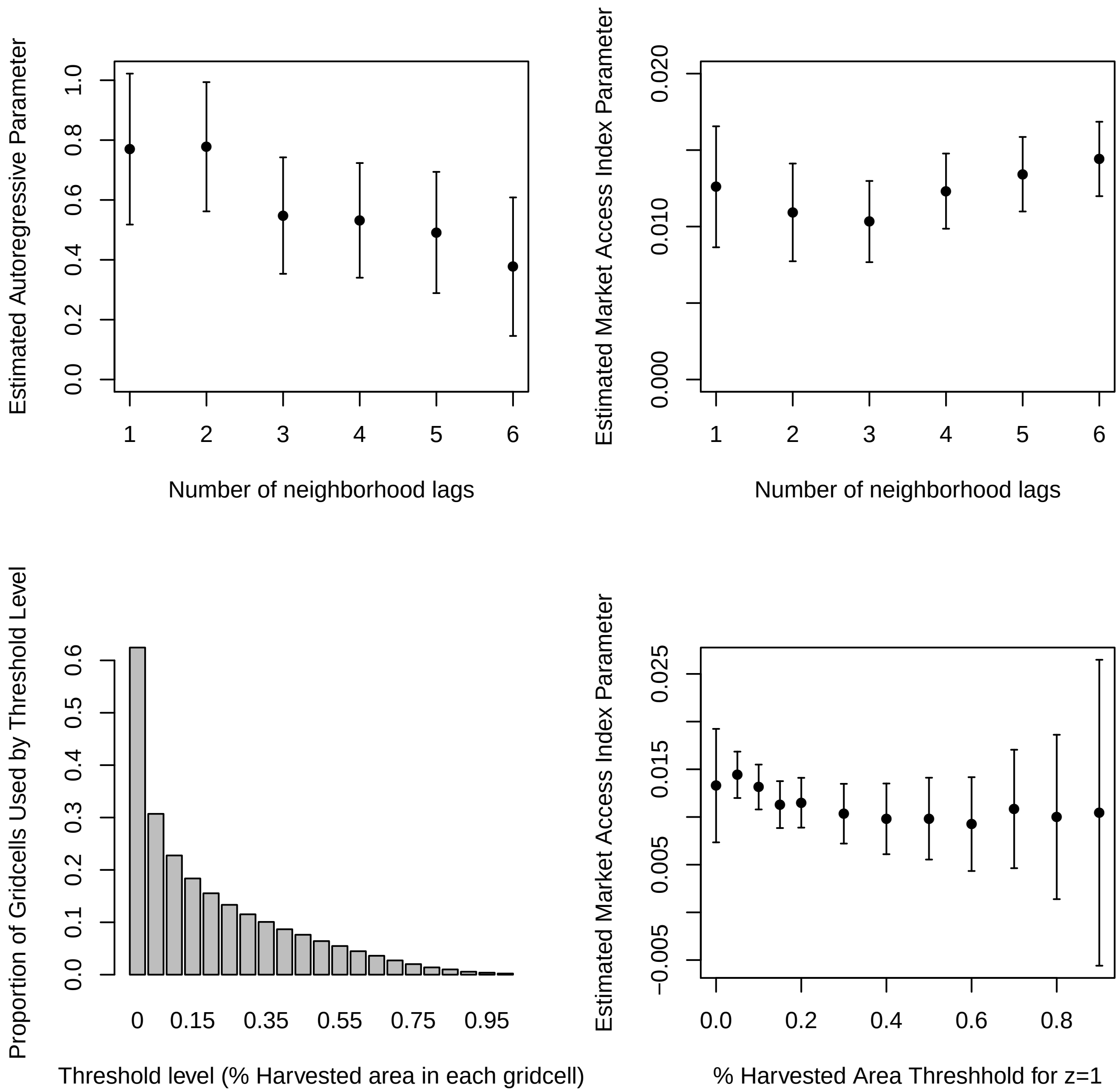

Figure 3: Robustness tests. (Error bars indicate a 95% confidence interval.) The upper left panel shows that the estimated autoregressive parameter decreases with the number of autoregressive lags; meanwhile, the estimate of the effect of market access (upper right panel) increases slightly as the correction for spatial autocorrelation improves. The lower left panel shows that the proportion of grid cells used in agriculture declines rapidly as the % of harvested area used to consider a given cell as used increases. Meanwhile, the lower right panel indicates that the range of variation for point-estimates of the market access parameter is relatively small, although the coefficient is less precisely estimated as we increase the threshold for use.

(2007) was identified as "managed" while the rest were considered "unmanaged." Notice that the implicit assumption made here is that managed lands are directly related to accessibility. The correspondence between accessibility and managed status is common in the economic literature on global land use change (Hertel et al., 2009) and is derived from the notion that a parcel of say, forest, will be commercially viable as long as the current value plus the expected value of the future stream of returns equals or surpasses the current cost of access (Sohngen et al., 2009).

With a clear division between accessible and inaccessible lands, we obtained a weighted-average elasticity of the land transition probability of each cover using as weights the predicted probabilities of land use. This produced an estimated non-cropland-to-cropland transition probability elasticity of 0.33 for forests and 0.23 for pastures. The comparable estimates by Lubowski (2002) are 0.34 and 0.30, respectively (table 1.) The remarkable closeness between these two sets of estimates obtained using very different data, samples and methods is reassuring.

Our next target for comparison is Scott (2013), who estimates the land supply elasticity from managed non-cropland to croplands in the U.S. using a dynamic model estimated with panel data; he obtains a point-estimate value of 0.3. In contrast to Lubowski (2002)'s, this elasticity refers to the percentage change in the number of hectares that are supplied to cropland given one percent in net returns to agriculture. In order to get a comparable figure, we used the elasticity of the land transition probabilities to changes in land rents to update the probability of land use given a 10% increase in land rents (equation 11), uniformly applied over the set of US accessible grid cells, both under forests and grasslands (at least conceptually, this matches Scott (2013)'s geographic domain.)

We then compared the updated predicted probability to the proportion of grid cells under cropland (that is, those identified as Z = 1 in the regression estimation), over the whole continent; in other words, τ in equation 12 equals 31% as per table 2. If the updated probability was greater than this proportion, we assumed that the entire grid cell converted into cropland. We expressed the new cropland as a percentage of the existing cropland and divided it by 10%, obtaining an elasticity of 0.26. As in the case of the elasticity of the land-use transition probabilities, this estimate is quantitatively indistinguishable from the estimates using times-series data and actual returns from Scott (2013).

The fact that our estimated land supply responses for the U.S. are very close to other estimates gives further confidence in our results, and encourages us to examine the diversity of land supply responses, which we do next in the context of an important policy issue: the indirect land use effects of biofuels.

Revisiting the Indirect Land Use Effects of Biofuels

We now explore the consequences of our findings for the analysis of Hertel et al. (2010) of the indirect land use effects of corn-ethanol expansion in the U.S. We chose to replicate this work because its indirect land-use estimates informed the biofuel regulations issued by the California Air

Resource Board (Babcock and Iqbal, 2014) and the data and model are freely available and readily accessible[2]. These authors model the expansion of US maize ethanol use from 2001 levels to the 2015 mandated level of 56.7 gigaliters by forcing 50.15 GL of additional ethanol production. As mentioned in the introduction, they calibrate global supply responses using the land use transition probabilities obtained from Lubowski (2002) for the U.S. Hertel et al. (2010) find that the biofuel mandate expands cropland by 1.6 Mha in the U.S. and by 2.6 Mha in the rest of the world. As per our replication, this expansion of cropland generates 953 teragrams of carbon dioxide (TgCO2).

Following Ahmed, Hertel, and Lubowski (2008), we use our estimates to calibrate the CET parameters of the land supply function for each region in Hertel et al. (2010) located in the Americas (see notes to table 4). The first step is to obtain the cross-price supply elasticities of forests and pastures by using equations 11-13 . Table 4 shows that these elasticities are quite different across regions. In the case of forests, a 1% increase in the relative profitability of agriculture reduces forest cover by 0.22% in the U.S. and only by between 0.01%- 0.03% in the rest of the continent. This result suggests that the access to the remaining forests in the continent is costly, and that high investments are needed to bring them into agricultural production.

Table 4: CET parameters for regions in Hertel et al.

	Revenue shares		Cropland-rent elasticity			
Region	Forests	Pastures	Forests	Pastures	Own	CET
Brazil	0.08	0.16	-0.01	-0.02	0.01	-0.02
Canada	0.59	0.06	-0.02	0.00	0.03	-0.05
USA	0.13	0.16	-0.22	-0.10	0.06	-0.22
Energy exporters	0.07	0.17	-0.03	-0.06	0.02	-0.06
Rest LA	0.14	0.12	-0.02	-0.05	0.01	-0.04

Notes: Energy exporters are Mexico, Colombia, Venezuela, and Argentina.The Rest of LA includes Chile, Uruguay, Paraguay, Bolivia, Peru, Ecuador, Panama, Costa Rica, Nicaragua, and Guatemala.

A similar pattern is observed for pastures and grasslands, although the gap between the U.S. and other regions is much smaller than in the case of forests. Table 4 also displays the region-specific own-price cropland elasticities and CET parameters (calibration details are in the appendix . Their pattern is inherited from the cross-price elasticities of the non-cropland uses. These results suggest that, although Latin America is one of the most land abundant regions in the world (Deininger and Byerlee, 2011), the changes in commodity prices needed to justify converting natural covers to agricultural production need to be large and sustained over time.

Figure 4 compares the original changes in land use and carbon dioxide emissions in Hertel et al. (2010) with the changes obtained by using our region-specific elasticities of transformation. A caveat to these results is that for the regions outside of the Americas, we kept the standard assumption of a single elasticity of land transformation as used in Hertel et al. (2010), which may also be difficult to justify in land abundant regions but that have high costs of land access, such as Africa (Chamberlin, Jayne, and Headey, 2014; Liu and Villoria, 2015).

[2]`https://www.gtap.agecon.purdue.edu/resources/res_display.asp?RecordID=3432`

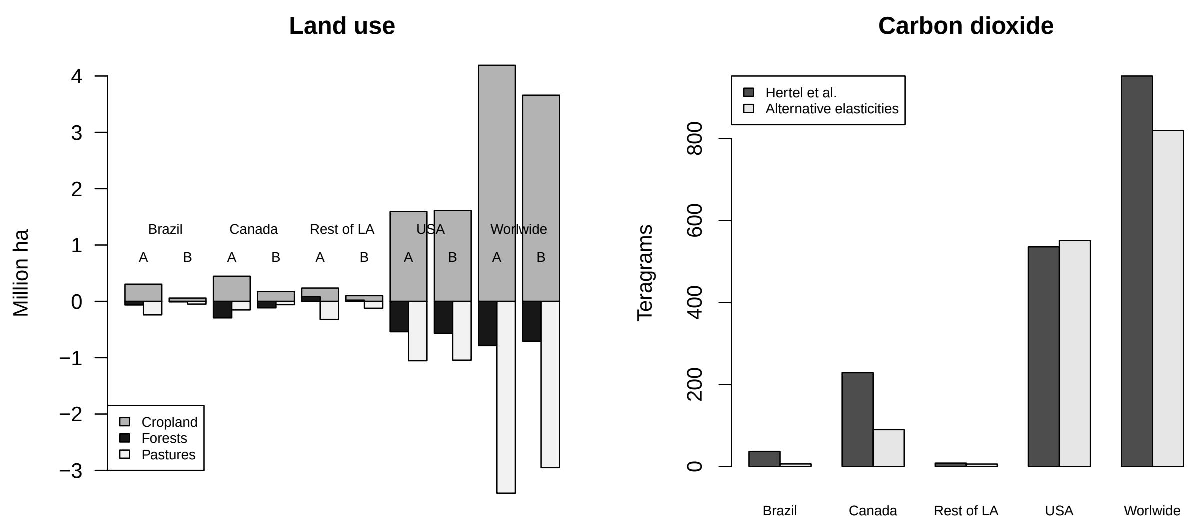

Figure 4: Revisiting the land use effects of biofuels from Hertel et al. (2010). The figure shows the global land conversion and associated greenhouse gas emissions due to increased maize ethanol production of 50.15 gigaliters per year starting in 2007. The original results in Hertel et al. (2010) are labeled as “A.” The results obtained using the elasticities obtained in this work (only for the U.S., Canada, Brazil, and the rest of Latin America) are labeled as “B.”

Few patterns are apparent from 4. Focusing first on the changes in land use—the left panel, where "A" identifies Hertel et al. (2010)'s results and "B" identifies our results—the estimates of land use change for the U.S. coincide across CET specifications. However, for Brazil, the region specific estimates predict around five times less cropland expansion than the original CET (304,759 ha vs 58,174 ha); for the rest of Latin America, the original CET calibration predicts cropland expansion of 145,283 ha above the region-specific estimates (100,839 ha); and for Canada, the results in Hertel et al. (2010) are 271,234 ha more than what we obtained with region-specific estimates. Notice too that by using region-specific estimates, we swap places between Brazil and the rest of Latin America, which now has a larger cropland response. This is in line with Babcock and Iqbal (2014)'s finding that model predictions from both GTAP and FAPRI/CARD of area expansion in Brazil are too high relative to the rest of South America. Relative to the region-specific elasticities, the imposition of a single CET across regions overestimates global cropland expansion by approximately 0.5 Mha and associated emissions by 133 TgCO2.

Conclusions

The geographic coverage of existing empirical estimates about the ease with which natural lands transition from natural uses into agriculture is insufficient to evaluate the global contributions of GHGE of a large number of policies concerned with food and environmental security. As Lubowski (2002), Rashford, Walker, and Bastian (2011), and Scott (2013) demonstrate, estimating land supply responses with explicit information on land transitions requires large amounts of very spatially and temporally detailed data on both land use choices and agricultural returns.

Our main contribution is to formulate a flexible empirical approach for analyzing land use responses based on an explicitly-spatial model of land-use choice. The chief advantage of this method is that it relies on readily available spatial data on land use patterns and biophysical attributes. This approach is much less onerous in terms of required data, particularly prices, which are generally unobservable at the land-plot level.

Another advantage of our approach is that identification of the land supply elasticities comes exclusively from changes in the physical determinant of the prices that determine land rents, while keeping land quality constant. This contrasts with methods using observed prices. For instance, Roberts and Schlenker (2013) acknowledge that their estimates for the U.S. may be capturing policies that incentivize setting land aside when crop prices are low. Likewise, Barr et al. (2011) document a large temporal variability of aggregated land supply for Brazil, which presumably captures policy decisions that affect the rate of land expansion/contraction, as documented by Malingreau, Eva, and de Miranda (2012).

The tractability of our approach comes with some costs. First, we do not know the input and output price elasticities with respect to market access, which could influence the effects of changes in agricultural net returns on the land transition probabilities. In addition, the underlying dependent variable is the percentage of each gridcell that is under cropland or improved pastures. By imposing a binary structure, we lose information that may be important; incorporating spatial effects on discrete choice models with explicit treatment of fractional data (Papke and Wooldridge, 2008) could help to remedy this caveat. Some limitations come from the data. In particular, land cover maps are notoriously uncertain (Fritz et al., 2011). The data we use is one of many competing products (Ramankutty and Foley, 1999; See et al., 2015). Therefore, investigating the sensitivity of our findings to different data sources could shed light on the degree to which uncertainties about land covers are reflected in the estimated land supply elasticities.

Despite these caveats, our results are robust to large variations in key assumptions underlying the estimation procedure. And our estimates for the U.S. are remarkably close to those in the existing literature. We take this as reassuring. The data on area changes after the high-price period from 2006 to 2010 suggests a more nuanced response that what would be predicted by land use models (Taheripour and Tyner, 2013; Babcock and Iqbal, 2014); our elasticities give a quantitative explanation as to why this may be the case. We revisit the estimates of indirect land use effects of US biofuels published by Hertel et al. (2010), and find that ignoring heterogeneous land supply responses leads to overestimating global cropland expansion and associated GHGE. Our findings underscore the importance of taking into account regional differences in global land supply responses.

References

Ahmed, S., T.W. Hertel, and R. Lubowski. 2008. "Calibration of a Land Cover Supply Function Using Transition Probabilities." GTAP Research Memorandum No. 14, GTAP Center, Department of Agricultural Economics, Purdue University.

Anselin, L. 2013. *Spatial Econometrics: Methods and Models*. Springer Science & Business Media.

—. 2002. "Under the Hood: Issues in the Specification and Interpretation of Spatial Regression Models." *Agricultural Economics* 27:247–267.

Avetisyan, M., U. Baldos, and T. Hertel. 2010. "Development of the GTAP Version 7 Land Use Data Base." *GTAP Research Memorandum* 19.

Babcock, B. 2009. "Measuring Unmeasurable Land-Use Changes from Biofuels." Iowa Ag Review No. 15, Iowa State University, Iowa.

Babcock, B., and Z. Iqbal. 2014. "Using Recent Land Use Changes to Validate Land Use Change Models." Staff Report No. 14 - SR 109, Iowa State University, Center for Agricultural and Rural Development, Ames, Iowa.

Barr, K.J., B.A. Babcock, M.A. Carriquiry, A.M. Nassar, and L. Harfuch. 2011. "Agricultural Land Elasticities in the United States and Brazil." *Applied Economic Perspectives and Policy* 33:449 –462.

Brady, M., and E. Irwin. 2011. "Accounting for Spatial Effects in Economic Models of Land Use: Recent Developments and Challenges Ahead." *Environmental and Resource Economics* 48:487–509.

Chamberlin, J., T.S. Jayne, and D. Headey. 2014. "Scarcity Amidst Abundance? Reassessing the Potential for Cropland Expansion in Africa." *Food Policy* 48:51–65.

Chomitz, K.M., and D.A. Gray. 1996. "Roads, Land Use, and Deforestation: A Spatial Model Applied to Belize." *The World Bank Economic Review* 10:487–512.

Cressie, N. 1993. *Statistics for Spatial Data*. New York, Chichester, Toronto, Brisbane, Singapore: John Wiley & Sons.

Deininger, K.W., and D. Byerlee. 2011. *Rising Global Interest in Farmland: Can it Yield Sustainable and Equitable Benefits?*. Washington D.C.: World Bank Publications.

Dumortier, J., D.J. Hayes, M. Carriquiry, F. Dong, X. Du, A. Elobeid, J.F. Fabiosa, and S. Tokgoz. 2011. "Sensitivity of Carbon Emission Estimates from Indirect Land-Use Change." *Applied Economic Perspectives and Policy*, Jul., pp. ppr015.

Eickhout, B., H. van Meijl, A. Tabeau, and T. van Rheenen. 2007. "Economic and Ecological Consequences of Four European Land Use Scenarios." *Land Use Policy* 24:562–575.

Fritz, S., L. See, I. McCallum, C. Schill, M. Obersteiner, M.v.d. Velde, H. Boettcher, P. Havlík, and F. Achard. 2011. "Highlighting Continued Uncertainty in Global Land Cover Maps for the User Community." *Environmental Research Letters* 6:044005.

Gohin, A. 2014. "Assessing the Land Use Changes and Greenhouse Gas Emissions of Biofuels: Elucidating the Crop Yield Effects." *Land Economics* 90:575–586.

Golub, A., T. Hertel, H. Lee, S. Rose, and B. Sohngen. 2009. "The Opportunity Cost of Land Use and the Global Potential for Greenhouse Gas Mitigation in Agriculture and Forestry." *Resource and Energy Economics* 31:299–319.

Greene, W.H. 2008. *Econometric Analysis*, 6th ed. Upper Saddle River NJ: Prentice Hall.

Gurgel, A., J.M. Reilly, and S. Paltsev. 2007. "Potential Land Use Implications of a Global Biofuels Industry." *Journal of Agricultural & Food Industrial Organization* 5.

Hayes, D., B. Babcock, J. Fabiosa, S. Tokgoz, A. Elobeid, T. Yu, F. Dong, C. Hart, E. Chavez, S. Pan, M. Carriquiry, and J. Dumortier. 2009. "Biofuels: Potential Production Capacity, Effects on Grain and Livestock Sectors, and Implications for Food Prices and Consumers." *Journal of Agricultural and Applied Economics* 41:465–491.

Hertel, T.W., A.A. Golub, A.D. Jones, M. O'Hare, R.J. Plevin, and D.M. Kammen. 2010. "Effects of US Maize Ethanol on Global Land Use and Greenhouse Gas Emissions: Estimating Market-Mediated Responses." *BioScience* 60:223–231.

Hertel, T.W., H. Lee, S. Rose, and B. Sohngen. 2009. "Modeling Land-Use Related Greenhouse Gas Sources and Sinks and Their Mitigation Potential." In T. W. Hertel, S. Rose, and R. Tol, eds. *Economic Analysis of Land Use in Global Climate Change Policy*. London and New York: Routledge, pp. 123–154.

Hertel, T.W., S. Rose, and R.S. Tol, eds. 2009. *Economic Analysis of Land Use in Global Climate Change Policy (Hardback) - Routledge*. Routledge Explorations in Environmental Economics, London and New York: Routledge, Taylor & Francis Group.

IGBP-DIS. 1998. "SoilData(V.0) A Program for Creating Global Soil-Property Databases, IGBP Global Soils Data Task, France."

IIASA/FAO. 2012. *Global Agro-Ecological Zones (GAEZ v3.0)*. IIASA, Laxenburg, Austria and FAO, Rome, Italy.

Jones, C.A., and R.D. Sands. 2013. "Impact of Agricultural Productivity Gains on Greenhouse Gas Emissions: A Global Analysis." *American Journal of Agricultural Economics* 95:1309–1316.

Keeney, R., and T.W. Hertel. 2009. "The Indirect Land Use Impacts of United States Biofuel Policies: The Importance of Acreage, Yield, and Bilateral Trade Responses." *American Journal of Agricultural Economics* 91:895–909.

Klier, T., and D.P. McMillen. 2008. "Clustering of Auto Supplier Plants in the United States." *Journal of Business & Economic Statistics* 26:460–471.

LeSage, J., and R.K. Pace. 2009. *Introduction to Spatial Econometrics*. CRC Press.

Liu, J., and N.B. Villoria. 2015. "Sub-Saharan Africa's Land Supply Response." In preparation. No. TBI, Purdue University, West Lafayette, IN, USA.

Lobell, D.B., U.L.C. Baldos, and T.W. Hertel. 2013. "Climate Adaptation as Mitigation: The Case of Agricultural Investments." *Environmental Research Letters* 8:015012.

Lubowski, R. 2002. "Determinants of Land-Use Transitions in the United States: Econometric Analysis of Changes among the Major Land-Use Categories." PhD dissertation, Harvard University.

Malingreau, J., H. Eva, and E. de Miranda. 2012. "Brazilian Amazon: A Significant Five Year Drop in Deforestation Rates But Figures are on the Rise Again." *AMBIO: A Journal of the Human Environment* 41:309–314.

McMillen, D.P. 2006. "Probit with Spatial Autocorrelation." *Journal of Regional Science* 32:335–348.

Miao, R., M. Khanna, and H. Huang. 2015. "Responsiveness of Crop Yield and Acreage to Prices and Climate." *American Journal of Agricultural Economics*, May, pp. aav025.

Monfreda, C., N. Ramankutty, and J.A. Foley. 2008. "Farming the Planet: 2. Geographic Distribution of Crop Areas, Yields, Physiological Types, and Net Primary Production in the Year 2000." *Global Biogeochemical Cycles*, Mar., pp. 1:19.

Monfreda, C., N. Ramankutty, and T. Hertel. 2009. "Global Agricultural Land Use Data for Climate Change Analysis." In T. W. Hertel, S. K. Rose, and R. S. Tol, eds. *Economic Analysis of Land Use in Global Climate Change Policy*. London and New York: Routdlege, pp. 33–49.

Nelson, G.C., V. Harris, and S.W. Stone. 2001. "Deforestation, Land Use, and Property Rights: Empirical Evidence from Darien, Panama." *Land Economics* 77:187–205.

New, M., M. Hulme, and P. Jones. 1999. "Representing Twentieth-Century Space–Time Climate Variability. Part I: Development of a 1961–90 Mean Monthly Terrestrial Climatology." *Journal of Climate* 12:829–856.

NOAA. 1995. "TerrainBase, Global 5 Arc-minute Ocean Depth and Land Elevation from the US National Geophysical Data Center (NGDC)."

Papke, L.E., and J.M. Wooldridge. 2008. "Panel Data Methods for Fractional Response Variables with an Application to Test Pass Rates." *Journal of Econometrics* 145:121–133.

Ramankutty, N., and J.A. Foley. 1999. "Estimating Historical Changes in Global Land Cover: Croplands from 1700 to 1992." *Global Biogeochemical Cycles* 13:PP. 997–1027.

Ramankutty, N., J.A. Foley, J. Norman, and K. McSweeney. 2002. "The Global Distribution of Cultivable Lands: Current Patterns and Sensitivity to Possible Climate Change." *Global Ecology and Biogeography* 11:377–392.

Rashford, B.S., J.A. Walker, and C.T. Bastian. 2011. "Economics of Grassland Conversion to Cropland in the Prairie Pothole Region." *Conservation Biology* 25:276–284.

Roberts, M.J., and W. Schlenker. 2013. "Identifying Supply and Demand Elasticities of Agricultural Commodities: Implications for the US Ethanol Mandate." *American Economic Review* 103:2265–95.

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* 95:395–406.

Scott, P.T. 2013. "Dynamic Discret Choice Estimation of Agricultural Land Use." Working paper, Toulouse School of Economics.

Searchinger, T., R. Heimlich, R.A. Houghton, F. Dong, A. Elobeid, J. Fabiosa, S. Tokgoz, D. Hayes, and T. Yu. 2008. "Use of U.S. Croplands for Biofuels Increases Greenhouse Gases Through Emissions from Land-Use Change." *Science* 319:1238–1240.

See, L., D. Schepaschenko, M. Lesiv, I. McCallum, S. Fritz, A. Comber, C. Perger, C. Schill, Y. Zhao, V. Maus, M.A. Siraj, F. Albrecht, A. Cipriani, M. Vakolyuk, A. Garcia, A.H. Rabia, K. Singha, A.A. Marcarini, T. Kattenborn, R. Hazarika, M. Schepaschenko, M. van der Velde, F. Kraxner, and M. Obersteiner. 2015. "Building a Hybrid Land Cover Map with Crowdsourcing and Geographically Weighted Regression." *ISPRS Journal of Photogrammetry and Remote Sensing* 103:48–56.

Siebert, S., J. Burke, J.M. Faures, K. Frenken, J. Hoogeveen, P. Döll, and F.T. Portmann. 2010. "Groundwater Use for Irrigation – a Global Inventory." *Hydrol. Earth Syst. Sci.* 14:1863–1880.

Smirnov, O.A. 2010. "Modeling Spatial Discrete Choice." *Regional Science and Urban Economics* 40:292–298.

Smith, P., M. Bustamante, H. Ahammad, H. Clark, H. Dong, E. Elsiddig, H. Haberl, R. Harper, J. House, M. Jafari, O. Masera, N. Ravindranath, C. Rice, C.R. Abad, A. Romanovskaya, F. Sperling, and F. Tubiello. 2014. "Agriculture, Forestry and Other Land Use (AFOLU)." In O. Edenhofer, R. Pichs-Madruga, Y. Sokona, E. Farahani, S. Kadner, K. Seyboth, A. Adler, I. Baum, S. Brunner, P. Eickemeier, B. Kriemann, J. Savolainen, S. Schlömer, C. von Stechow, T. Zwickel, and J. Minx, eds. *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY, USA.: Cambridge University Press.

Sohngen, B., C. Tennity, M. Hnytka, and K. Meeusen. 2009. "Global Forestry Data for the Economic Modeling of Land Use." In T. W. Hertel, S. K. Rose, and R. S. Tol, eds. *Economic Analysis of Land Use in Global Climate Change Policy*. West Lafayette, IN: Center for Global Trade Analysis, Department of Agricultural Economics, Purdue University, pp. 49–72.

Stevenson, J.R., N. Villoria, D. Byerlee, T. Kelley, and M. Maredia. 2013. "Green Revolution Research Saved an Estimated 18 to 27 Million Hectares from Being Brought into Agricultural Production." *Proceedings of the National Academy of Sciences* 110:8363–8368.

Taheripour, F., and W. Tyner. 2013. "Biofuels and Land Use Change: Applying Recent Evidence to Model Estimates." *Applied Sciences* 3:14–38.

Tubiello, F.N., M. Salvatore, R. Cóndor-Golec, A. Ferrara, S. Rossi, R. Biancalani, S. Federici, H. Jacobs, and A. Flammini. 2014. "Agriculture, Forestry and Other Land Use Emissions by Sources and Removals by Sinks." ESS No. 14-02, Food and Agriculture Organisation of the UN, Rome, Italy.

Ullman, E. 1941. "A Theory of Location for Cities." *American Journal of Sociology* 46:853–864.

van Meijl, H., T. van Rheenen, A. Tabeau, and B. Eickhout. 2006. "The Impact of Different Policy Environments on Agricultural Land Use in Europe." *Agriculture, Ecosystems & Environment* 114:21–38.

van Velthuizen, H., B. Huddleston, G. Fischer, M. Salvatore, E. Ataman, F. Nacthergaele, M. Zanetti, and M. Bloise. 2006. "Mapping Biophysical Factors that Influence Agricultural Production and Rural Vulnerability." Working paper No. 11, Food and Agriculture Organisation of the United Nations & IIASA.

Verburg, P.H., E.C. Ellis, and A. Letourneau. 2011. "A Global Assessment of Market Accessibility and Market Influence for Global Environmental Change Studies." *Environmental Research Letters* 6:034019.

Verburg, R., E. Stehfest, G. Woltjer, and B. Eickhout. 2009. "The effect of Agricultural Trade Liberalisation on Land-Use Related Greenhouse Gas Emissions." *Global Environmental Change* 19:434–446.

Villoria, N.B., D. Byerlee, and J. Stevenson. 2014. "The Effects of Agricultural Technological Progress on Deforestation: What Do We Really Know?" *Applied Economic Perspectives and Policy* 36:211–237.

Villoria, N.B., and T.W. Hertel. 2011. "Geography Matters: International Trade Patterns and the Indirect Land Use Effects of Biofuels." *American Journal of Agricultural Economics* 93:919–935.

Appendices

Table A-1: Data sources.

Variables	Units[Resolution(min)*]	Source
Potential vegetation	5 types[a][5]	Ramankutty and Foley (1999)[c]
Soil fertility constraints	0-8[a][5]	IIASA/FAO (2012)[d]
Precipitation	mm/year[30]	New, Hulme, and Jones (1999)[c]
Elevation	meters[5]	NOAA (1995)[c]
Soil pH	0-14[30]	IGBP-DIS (1998)[c]
Soil carbon (1m depth)	kg-C/m^2[30]	IGBP-DIS (1998)[c]
Temperature[b]	^{o}C[30]	New, Hulme, and Jones (1999)[c]
Market access	Index[5]	Verburg, Ellis, and Letourneau (2011)[e]
Area equipped for irrigation	% of gridcell[5]	Siebert et al. (2010)[f]
Built-up land	% of gridcell[5]	Miteva, B.[c]
Protected areas	0-1[a][5]	van Velthuizen et al. (2006)[g]
GTAP Agro-ecological zones	AEZs[5]	Monfreda, Ramankutty, and Hertel (2009)[h]

* Datasets available at a 30-min resolution were resampled to 5-min by assuming that all the 5-min degree cells within a 30-min degree cell have the same value.

[a] Description of categories for (i) Potential vegetation: Original 15 categories reduced to five categories as displayed in table A-2; (ii) Soil fertility constraints categories: (0) undefined, (1) no constraints, (2) slight constraints, (3) moderate, (4) constrained, (5) severe, (6) very severe, (7) unsuitable for cultivation, (8) water; (iii) Protected areas: (Original categories were 0) non-protected areas, (1) protected areas where agriculture should not be occurring, and (2) protected areas where agriculture could not be occurring; we coded them as a dummy that equals one for protected areas and zero otherwise.

[b] Based on monthly average temperatures (1961-1990), we create a growing season average based on hemispheres:

Season	North	South
Winter	Dec-Jan-Feb	Jun-Jul-Aug
Spring	Mar-Apr-May	Sep-Oct-Nov
Summer	Jun-Jul-Aug	Dec-Jan-Feb
Autumn	Sep-Oct-Nov	Mar-Apr-May

[c] Data downloaded from the Atlas of the Biosphere, an initiative of the Center for Sustainability and the Global Environment (SAGE) at the University of Wisconsin, `http://www.sage.wisc.edu/atlas/`, last accessed on May 26, 2015. The Atlas strives to "include only those datasets that presumably represent the pinnacle of current knowledge in a particular field."

[d] Data downloaded from `http://www.iiasa.ac.at/Research/LUC/GAEZ/plt/zip/plate22.zip`,

last accessed on May 26, 2015.

[e]Data downloaded from `http://www.ivm.vu.nl/en/Images/marketinfluence_tcm53-229510.zip`, last accessed on May 26, 2015.

[f]Data downloaded from `http://www.fao.org/nr/water/aquastat/irrigationmap/index10.stm`, last accessed on May 26, 2015.

[g]Data downloaded from `http://www.fao.org/geonetwork/srv/en/metadata.show?id=31040`, last accessed on May 26, 2015.

[h]Data downloaded from `https://www.gtap.agecon.purdue.edu/resources/res_display.asp?RecordID=3184`, last accessed on May 26, 2015.

Table A-2: Correspondence between land cover types used in this work and the potential vegetation types of Ramankutty and Foley (1999) .

Land covers (this work)	Potential Vegetation (Ramankutty and Foley, 1999)
Tropical forests	(1) Tropical Evergreen Forest/Woodland
	(2) Tropical Deciduous Forest/Woodland
Temperate forests	(3) Temperate Broadleaf Evergreen Forest/Woodland
	(4) Temperate Needleleaf Evergreen Forest/Woodland
	(5) Temperate Deciduous Forest/Woodland
	(6) Boreal Evergreen Forest/Woodland
	(7) Boreal Deciduous Forest/Woodland
	(8) Mixed Forest
Savannas& Grasslands	(9) Savanna
	(10) Grassland/Steppe
Shrublands	(11) Dense Shrubland
	(12) Open Shrubland
Other	(13) Tundra
	(14) Desert
	(15) Polar Desert/Rock/Ice

Calibrating a Constant Elasticity of Transformation Function Using Land-use Transition Probabilities

1. Denote cropland by agr and non-croplands by $nagr$, i.e., pastures and forests. Let R denote net returns to agriculture. Furthermore, let θ denote land-use revenue shares so that $\theta_{agr} + \sum_{nagr} \theta_{nagr} = 1$. All the variables in this algorithm are region-specific; that is, they apply to a defined set of grid-cells.

2. We estimate cross-price elasticities for pastures and forests, $\eta_{nagr,agr} = \dfrac{\delta nagr}{nagr}\dfrac{R}{\delta R}$, using

equations 11-13.

3. Ahmed, Hertel, and Lubowski (2008) demonstrate that the CET parameter of the land supply function, σ, is given by:
$$\sigma = -\frac{\eta_{agr,agr}}{1-\theta_{agr}}. \tag{A-1}$$

4. By virtue of the homogeneity of supply, we obtain:
$$\eta_{agr,agr} + \sum_{agr} \eta_{agr,nagr} = 0. \tag{A-2}$$

5. The task then is to recover $\eta_{agr,nagr}$ from our estimates of $\eta_{nagr,agr}$.

6. For this, take advantage of the symmetry of the matrix of Allen-partial elasticities of substitution, in particular, for any non-cropland use $nagr$:
$$\sigma_{nagr,agr} = \frac{\eta_{nagr,agr}}{\theta_{agr}} = \sigma_{agr,nagr} = \frac{\eta_{agr,nagr}}{\theta_{nagr}}. \tag{A-3}$$
from which we can readily recover the $\eta_{agr,nagr}$, plug them into equation A-2, and solve for $\eta_{agr,agr}$, which yields σ by virtue of equation A-1.

7. The land use revenue shares for each region in Hertel et al. (2010) are displayed in table 4; these are taken from the GTAP land use database V.7 (Avetisyan, Baldos, and Hertel, 2010, last accessed June 16, 2015) which is freely downloaded from `https://www.gtap.agecon.purdue.edu/resources/res_display.asp?RecordID=3426`.

Supporting Online Materials

Sample Description and Representativeness vis a vis Population

Table S-1: Proportion of gridcells used in agriculture in each land cover type (population vs. sample)

	UsedInData	UsedInSample
A	0.03	0.03
Ft	0.18	0.19
FT	0.27	0.27
G	0.51	0.51
S	0.27	0.26

Table S-2: Share of each land cover type in natural (Z=0) and agricultural (Z=1) land use (population vs. sample)

	0	1	0	1
A	0.04	0.00	0.04	0.00
Ft	0.33	0.17	0.33	0.17
FT	0.30	0.24	0.29	0.25
G	0.21	0.48	0.21	0.48
S	0.13	0.11	0.14	0.11
Sum	1.00	1.00	1.00	1.00

Table S-3: Proportion of gridcells used in agriculture in each soil fertility category (1 is most fertile, population vs. sample)

	UsedInData	UsedInSample
1	0.57	0.57
2	0.39	0.40
3	0.35	0.34
4	0.40	0.40
5	0.23	0.24
6	0.12	0.11
7	0.13	0.14

Table S-4: Share of each soil fertility category (1 is most fertile) in aggregated natural (Z=0) and agricultural (Z=1) land use (population vs. sample)

	0	1	0	1
1	0.12	0.36	0.12	0.35
2	0.07	0.11	0.08	0.11
3	0.08	0.10	0.08	0.09
4	0.14	0.20	0.13	0.20
5	0.13	0.09	0.13	0.09
6	0.24	0.07	0.24	0.07
7	0.22	0.08	0.22	0.08
Sum	1.00	1.00	1.00	1.00

Table S-5: Proportion of gridcells under agricultural use in each country (population vs. sample)

	UsedInData	UsedInSample
ARG	0.28	0.28
BOL	0.16	0.16
BRA	0.24	0.24
CAN	0.20	0.20
CHL	0.12	0.13
COL	0.23	0.21
CRI	0.40	0.38
ECU	0.48	0.44
GTM	0.55	0.54
GUF	0.00	0.00
GUY	0.04	0.05
HND	0.54	0.56
MEX	0.38	0.38
NIC	0.45	0.46
PAN	0.28	0.33
PER	0.14	0.14
PRY	0.31	0.31
SLV	0.96	1.00
SUR	0.01	0.02
URY	0.25	0.22
USA	0.51	0.51
VEN	0.12	0.12

Table S-6: Proportion of gridcells under agricultural use in each AEZ (population vs. sample)

	UsedInData	UsedInSample
1	0.17	0.18
2	0.31	0.33
3	0.43	0.44
4	0.38	0.36
5	0.24	0.25
6	0.12	0.12
7	0.24	0.23
8	0.51	0.50
9	0.38	0.38
10	0.63	0.64
11	0.72	0.71
12	0.55	0.54
13	0.17	0.19
14	0.05	0.05
15	0.08	0.08
16	0.17	0.15
17	0.00	0.00
18	0.00	0.00

Table S-7: Average market access of lands in natural and agricultural use (population vs. sample)

Z	x	InSample
0.00	6.67	6.67
1.00	22.70	22.91

Table S-8: Average market access of each soil fertility category in (population vs. sample)

SoilFert	x	InSample
1.00	15.56	15.52
2.00	16.02	16.38
3.00	15.43	15.45
4.00	18.60	18.76
5.00	9.63	9.75
6.00	5.51	5.68
7.00	4.79	4.87

Table S-9: Average market access of each land cover type (population vs. sample)

cover	x	InSample
A	4.55	4.10
Ft	5.40	5.58
FT	16.80	16.93
G	12.54	12.70
S	13.05	12.56

Table S-10: Proportion of gridcells in each soil fertility-land cover type combination (sample)

	A	Ft	FT	G	S
1	0.10	0.02	0.06	0.41	0.38
2	0.11	0.03	0.07	0.12	0.16
3	0.10	0.07	0.08	0.08	0.13
4	0.10	0.18	0.18	0.13	0.13
5	0.13	0.18	0.10	0.08	0.10
6	0.12	0.41	0.16	0.08	0.05
7	0.34	0.11	0.36	0.11	0.06
Sum	1.00	1.00	1.00	1.00	1.00

Table S-11: Average market access of each AEZ (population vs. sample)

AEZ18	x	InSample
1	7.33	6.55
2	10.68	11.44
3	9.67	8.56
4	11.07	11.07
5	6.23	6.47
6	3.90	4.10
7	9.32	9.04
8	9.88	10.21
9	12.55	12.29
10	32.58	32.87
11	38.48	38.71
12	22.83	22.23
13	4.70	4.41
14	2.54	2.64
15	2.63	2.79
16	6.84	6.56
17	0.00	0.00
18	0.00	0.00

Table S-12: Average market access of gridcells in agriculture in each country (population vs. sample)

Country	x	InSample
ARG	14.56	15.17
BOL	7.69	7.24
BRA	13.30	13.28
CAN	20.84	21.26
CHL	27.43	25.63
COL	28.76	30.61
CRI	24.75	25.93
ECU	10.78	12.04
GTM	40.65	39.97
GUY	9.18	8.64
HND	39.35	37.47
MEX	26.66	26.54
NIC	35.23	33.02
PAN	6.57	8.38
PER	4.94	6.00
PRY	6.89	7.22
SLV	66.05	68.61
SUR	21.14	0.00
URY	13.67	10.09
USA	28.78	28.82
VEN	23.48	27.47