Selected Presentation at the 2020 Agricultural & Applied Economics Association Annual Meeting, Kansas City, Missouri, July 26-28

Copyright 2020 by authors. All rights reserved.
Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

Estimating the spatially heterogeneous elasticities of land supply to U.S. crop agriculture

Shourish Chakravarty[1], Nelson B. Villoria[2]

Introduction

The biggest issue that US agriculture is currently facing is its long term sustainability given the increased stress on land and water resources resulting from increasing agricultural production. For example, intensification of production has led to large amounts of nitrogen being discharged into surface and groundwater (Goolsby et al., 2001; Turner and Rabalias, 1991). Elevated nitrogen levels in waterways are detrimental to plant and animal life, agriculture and human health as well. Along with water quality, water quantity poses a threat to agricultural sustainability. In addition to water resources, expanding croplands have put pressure on land with conversion of environmentally sensitive and erosive lands to agriculture being the biggest concern (Lark et al., 2015). Moreover, the effectiveness of Conservation Reserve Program in achieving its objective of keeping erosive land out of agriculture has been restricted by commodity prices and land returns (Hellerstein and Malcolm, 2010). While various models have been used to assess long run sustainability of US agriculture, an important limitation of the existing research is the limited spatial resolution of the global to national scale modelling. This restricts their usage in understanding local environmental impacts and sustainability stresses since land supply responses vary by locality due to agro-ecological, economic and institutional factors. The key indicator of land supply response is the land supply elasticity which is the percentage change in cropland due to a percentage change in land rents accruing to agriculture. The primary objective of this paper is to estimate geographically explicit land supply elasticities to changes in agricultural cash rents, at a resolution of 5 arc minutes or between 5500 and 7600 hectares depending on the latitude, for the contiguous US using panel data on land use and controlling for land quality attributes.

Land supply elasticities determine the acreage of natural lands that get converted into croplands when agricultural productivity increases. Thus, models that predict environmental indicators related to land conversion, like greenhouse gas emissions, biodiversity loss, or changes in water usage, could condition on spatially explicit land supply elasticities to more precisely estimate the location of the changes in the indicators (Villoria and Liu, 2018). Villoria and Liu (2018) estimated the spatial pattern of land supply elasticities with respect to changes in market access for North and South America using cross sectional data. The limitation of those estimates is that they need to be scaled to a known elasticity, which are unobserved for most of the countries in the study. However, the spatially heterogeneous elasticities that the authors estimate, reasonably predict the regions where cropland conversions primarily took place as evidenced by estimates found by Lark et al. (2015) and Graesser et al. (2015). In this paper we estimate more accurate land supply elasticities for the US by using time series data on county level cash rents from the USDA National Agricultural Statistical Service (NASS) and patterns of land cover change from NASS' Cropland Data Layer (CDL). In addition to estimating high resolution land supply elasticities, our research will also contribute to improving the ability of economic models to assess impacts of agricultural policies.

[1] Postdoctoral Scholar, Department of Forestry and Natural Resources, Purdue University
[2] Associate Professor, Department of Agricultural Economics, Kansas State University

The gridcell level elasticities will also be used in a SIMPLE-G framework, which is a multi-region, partial equilibrium model of gridded cropland use, crop production, consumption and trade (Baldos and Hertel, 2013), to assess policies aimed at improving agricultural sustainability.

Conceptual framework

We assume that there are two land uses – cropland and non-cropland. For a landowner, the utility from converting a piece of land i from its natural state to cropland at time t, U_{it} can be represented by a probit model of binary choice,

$$Prob(I = 1) = \Phi(U_i). \qquad (1)$$

The indicator I is equal to 1 when parcel i is a cropland and 0 otherwise, and Φ denotes the standard normal cumulative distribution function (CDF). U_{it} can be specified as a function of county level land rents accruing to agriculture R_{ct}, where *c* is a county index, grid cell level access index A_i (Verburg et al, 2011), and grid cell level soil quality index Q_i, ranging from 0 to 1, as represented by the National Commodity Crop Productivity Index (NCCPI) which is derived from the Soil Survey Geographic (gSSURGO, 2017) database maintained by USDA.

$$U_{it} = \beta_0 + \beta_1 R_{ct} + \beta_2 A_i + \beta_3 Q_i + \varepsilon_{it} \qquad (2)$$

The parameters from the utility function above are estimated using the regression model of the form:

$$Z_{ict} = \Phi(\beta_0 + \beta_1 R_{ct} + \beta_2 A_i + \beta_3 Q_i + \varepsilon_{it}) \qquad (3)$$

where Z_{ict} is the fraction of each grid cell under cropland. Z_{ict} is obtained from the Cropland Data Layer (CDL) data from USDA which classifies land parcels at 30m resolution into over 120 different land use types. We reclassify the land use types into five broad land use categories – cropland, forest, grassland and pasture, shrubland, and urban. Next, we aggregate the small 30m-by-30m pixels of land use to 5 arc minute level grid cells and calculate the proportions of each of the five broad land use classification type per grid cell. Given the fractional nature of Z_{ict}, we estimate the above specification using a fractional probit model for panel data as outlined by Papke and Wooldridge (2008). The main advantage of using a fractional probit instead of a logistic model in our study is that the conditional logit MLE is not consistent unless the dependent variable is binary in nature (Papke and Wooldridge, 2008). Another important issue in the estimation of the fractional probit model is the spatial autocorrelation of ε_{it}. We plan to follow Villoria and Liu's (2018) four pronged strategy to control for spatial autocorrelation. The strategy includes (i) using a spatial bootstrap resampling algorithm proposed by Zhu and Morgan (2004) to obtain empirical variance estimators robust to spatial autocorrelation; (ii) including spatial lags of the independent variables; (iii) estimate the model using random samples of approximately 10% of the data using a sampling scheme that preserves the gridded structure of the original data; and (iv) capture the uncertainty from sampling by using the resampling techniques proposed by Kleiner et al. (2014).

After estimating the model, we estimate the grid cell specific elasticities of the change in cropland to change in land rents by

$$\epsilon_i = \frac{\delta \hat{Z}_i}{\delta R_c} \times \frac{R_c}{\hat{Z}_i} = \phi[\hat{\beta}_0 + \hat{\beta}_1 R_{ct} + \hat{\beta}_2 A_i + \hat{\beta}_3 Q_i]\hat{\beta}_1 \times \frac{R_c}{\hat{Z}_i} \qquad (4)$$

where ϕ is the standard normal density function and $\widehat{Z_\iota}$ are fitted cropland shares computed using parameter estimates from the fractional probit model. The partial effects, $\frac{\delta \widehat{Z_\iota}}{\delta R_c}$ are specific to each grid cell.

Data

In this paper we have used a variety of geo-spatial datasets of varying resolutions which have been aggregated or resampled to have a uniform resolution of 5 arcminutes. Table 1 lists the variables used, their original resolution, and the sources from where the relevant datasets and in turn the variables are obtained.

Table 1: Data Sources

Variables	Units [resolution/admin level]	Source	Years
Cropland Data Layer (CDL)	Type (30 m)	USDA-NASS	2009 – 2017
Cropland area	County	Agricultural Census, USDA	2007, 2012, 2017
Market Access	Index (5 arcminute)	Verburg, Ellis, Letourneau (2011)	2011
Soil Quality	Index (10 m)	gSSURGO - USDA	Jan 2019 release
Cash Rents	County	USDA-NASS	2009-14, 2016-17
CRP Payments	County	USDA-NASS	2009-17
Farm Resource Regions	Aggregated county	USDA-ERS	NA
Level III Eco-regions	85 eco-regions	EPA	NA
Irrigation	250 m	USGS (MODIS MIrAD-US)	2007, 2012, 2017
Precipitation	mm year (county/5 arcmin)	PRISM	2009 – 2017
Growing Degree Days	degreedays (county/5 arcmin)	PRISM	2009 – 2017
Cooling Degree Days	degreedays (county/5 arcmin)	PRISM	2009 – 2017
Bio-diesel and ethanol plants	# plants	National Biorefineries Database	2009
Population (yearly)	County	Census	2009 – 2017
Urban area	Area/ region	Census TIGER/Line shapefile	2019

The unit of analysis in this paper is a 5 arcminute gridcell for contiguous US. The 5 arcminute gridcells are inclusive of the gridcells that are used in the SIMPLE-G partial equilibrium model developed by Baldos and Hertel (2012). Since the gridcells are measured in arcminutes, we use a uniform geographic coordinate system (GCS) WGS-1984 for all the geospatial datasets used in our analysis. The gridcells from the SIMPLE-G framework serve as the reference to which the gridcells from aggregated and resampled datasets like the Cropland Data Layer are aligned. We use Cropland Data Layer (CDL) for creating a time series (yearly) data of proportion of area classified as cropland in each of the 5 arcminute gridcells for the years 2009 through 2017. The CDL data are obtained from USDA's National Agricultural Statistics Service (NASS) and are in the form of raster files with Albers equal-area conic projection system and a resolution of 30m. The CDL classifies crops and vegetation into more than 130 categories including over 100 crop categories. These categories are created by NASS using supervised classification of pixels from Landsat satellite images.

We reclassify all crop categories and fallow or idle cropland into a single cropland classification and create a new raster data (Lark et al., 2017). Next, we change the projection of the reclassified cropland raster to GCS WGS-1984. We aggregate the reprojected rasters to the closest integer value of the ratio of the gridcell size of the 5 arc min reference raster to that of the reprojected cropland rasters. The ratio is 239.810, and therefore we use 240 as the cell size. The aggregated reprojected cropland rasters are next

resampled to have the same sized gridcells as the reference raster. However, in order to have the gridcells of the cropland rasters aligned to those in the reference rasters, in the Resample tool of ArcGIS, the Processing Extent option in the Environment Settings is selected, and the Extent and the Snap Raster are changed to be the same as the reference 5 arcminute raster. This ensures that the centroids and thus the corners of the gridcells for both the cropland and the reference rasters coincide. After completing the procedures in ArcGIS, for each of the new aggregated 5 arcminute resolution cropland raster, the raster is converted to XYZ (or XY coordinate and Z data) points in R and saved as csv. The coordinates of the centroids of the gridcells match to those of the reference raster, and we follow the same procedure for obtaining 5 arcminute gridcell level data for area classified as non-cropland. The four primary non-cropland land use/cover types in our analysis are forests, grassland and pastures, shrubland, and urban builtup. Unlike the extensive cropland classification, CDL does not have numerous classifications for forests or other primary land use/cover types.

The gridcell level data for cropland and other land cover types do not have information on which counties and states they belong to. In order to find the exact counties for the gridcells, we use the Identity tool in ArcGIS. The Identity tool "Computes a geometric intersection of the input features and identity features. The input features or portions thereof that overlap identity features will get the attributes of those identity features" (ESRI). We use contiguous US county shapefile as the file with identity features, and one of the 5 arcminute cropland rasters as the file with input features and use to tool to find corresponding counties for each of the gridcells. Post matching the gridcells with counties, we create a unique ID for each of the gridcell and create a data key with information on gridcells, the corresponding coordinates of the centroids, state and counties, and their IDs. This key is then used to match grdicells from all the 5 arcminute rasters to the respective counties. Representative maps for four primary land use/cover classification types from 2016 CDL are given in figure 1. The blue/green regions correspond to each land use/cover type

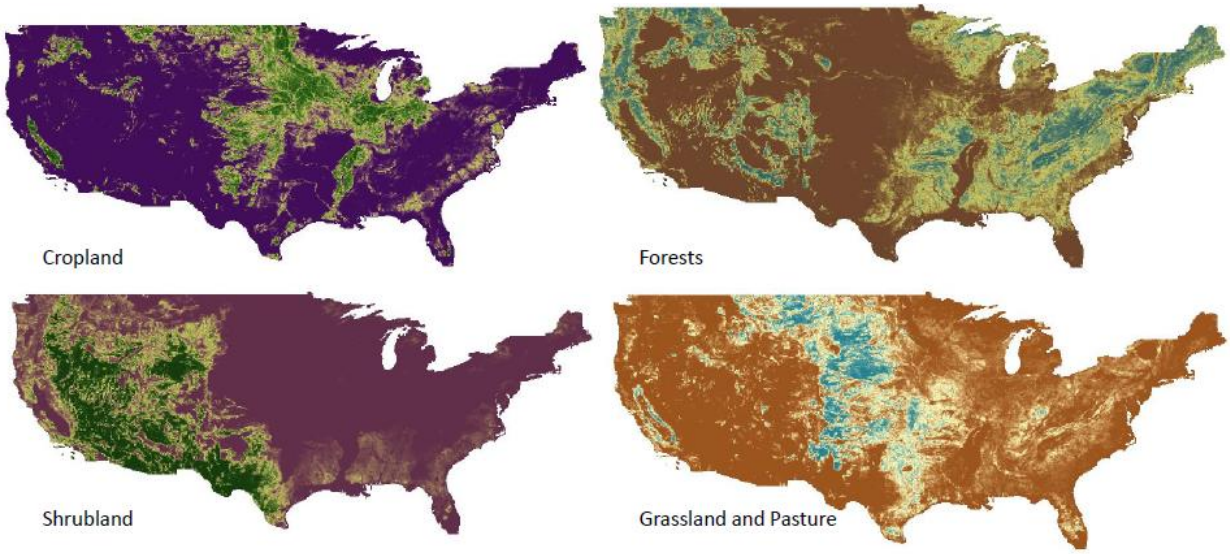

Figure 1: Primary land use/cover types from CDL 2016

In this paper, we consider cropland data from 2009 to 2017. This is primarily because, as will be discussed later, county level cash rent data from USDA-NASS are missing for most of the counties in 2008. Table 2 gives brief descriptive statistics for the five primary land use/cover types for all the years together. The

mean refers to the proportion of contiguous US that is classified as a particular land use/cover type. Forests are the most common land cover type, followed by shrublands, grassland and pasture, and cropland. The maps in figure 1 show the distinct pattern of land use/cover type in the US.

Table 2: Descriptive Statistics - Primary land classification types

Classification	#gridcells per year	Mean	Median	s.d.	Min	Max
Cropland	119,756	0.163	0.012	0.256	0	1
Forest	119,756	0.268	0.122	0.304	0	1
Grassland & Pasture	119,756	0.206	0.1	0.252	0	1
Shrubland	119,756	0.221	0.032	0.328	0	1
Urban	119,756	0.057	0.032	0.112	0	1

Between 2009 and 2017, cropland area has increased in the US. There is significant geographic heterogeneity in terms of where changes have taken place. For example, between 2008 and 2012 conversion to agriculture has primarily taken place in the plains states (Lark et al., 2015). Thus, grassland to cropland conversion was more common than conversion from other land cover types to cropland. In order to control for the effects of non-cropland dominant land use/cover type for gridcells on land supply elasticity, we create dummies for non-cropland dominant land use/cover type. Forest is the dominant non-cropland land use/cover type, followed by grassland and shrubland. For any gridcell, the dominant land cover type could be marginally more than the next most dominant land cover type. In order to control for this, we compare land covers of gridcells such that if a land cover proportion falls within a +-10% range of the next highest land cover proportion, the two land cover proportions are considered to be the same. Thus, in addition to forest, grassland and shrubland land cover types, we also have FG (forest and grassland), FS (forest and shrubland), SG (shrubland and grassland), and FGS (forest, grassland, and shrubland) and the dominant non-cropland land cover types. However, the number of cells in the latter categories is not high. Since a very small proportion of gridcells have dominant urban land cover type, we do not consider them in figure 2.

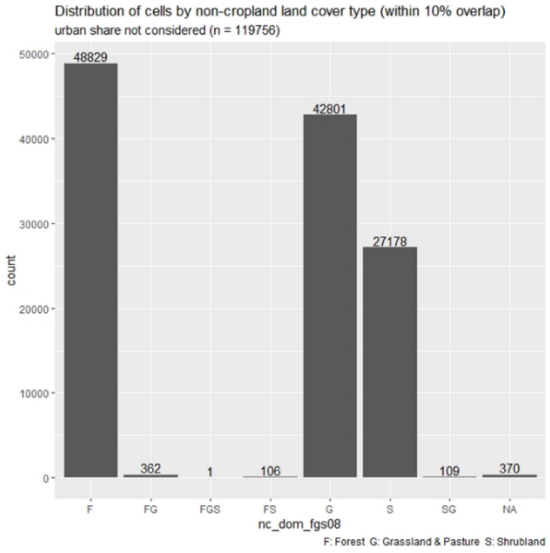

Figure 2: Number of gridcells by dominant non-cropland land cover type

We compare regions which saw land conversion as calculated from CDL to those estimated by Lark et al., (2015). Level III eco-regions of the US are used to recognize the regions where cropland expansion took place. Eco-regions refer to regions where ecosystems are similar. There are 85 eco-regions for contiguous US. Eco-region shapefiles are obtained from EPA and then merged to the gridcell key to match gridcells to their respective eco-regions. We take the average of cropland proportion for the years 2008, 2009 and 2010 and compare then to the average of cropland proportion for the years 2012, 2013 and 2014. The difference in average cropland proportion by eco-region is given in figure 3. Similar to Lark et al. (2015) we find that the Northern Great Plains and the High Plains regions had amongst the highest proportions of land converted to agricultural land. Moreover, eco-regions that had higher proportion of land classified as cropland at the baseline had higher expansion in cropland. The trends are similar for period between 2016-18 and 2008-10.

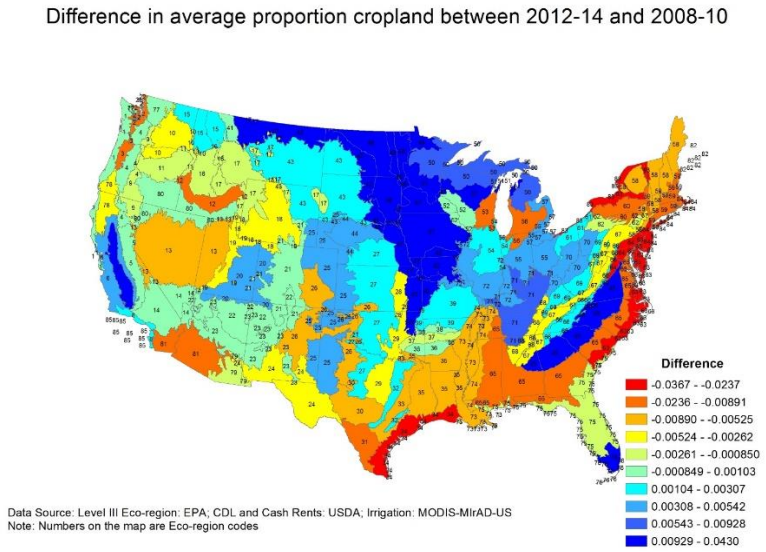

Figure 3: Cropland expansion by eco-region

CDL uses Landsat imagery to classify vegetation and crops. CDL data are generated by supervised classification of images which involves first manually recognizing pixels from images associated with particular crop or vegetation type, and then using the recognized sites (training sites) to analyze all other pixels in the images of the crops (USGS). Thus, CDL differs from data based on surveys like the Agricultural Census. Lark et al. (2017) outlines challenges using CDL data and recommends practices to improve accuracy of the data used. The key challenges of using CDL data and how we have addressed them in this paper are as follows:

1. Using data for specific crops could lead to inaccurate estimation of acreage under those crops. We follow Lark et al. (2017)'s recommendation of combining all crop classes to create a single combined cropland category.
2. Using CDL data for only specific years could lead to incorporating any misclassification that could have occurred for a particular year. Instead we utilize most of the temporal data in our panel data analysis. We also take 3 year averages of cropland data when we estimate

elasticities using long differences estimation. For example, we take the average of cropland proportion for 2008-10 as an approximate for the year 2009.

3. Verifying with independent data improves accuracy of estimates. We compare our cropland estimates to those from Agricultural Census. For most of the states, total cropland from Agricultural Census is consistently 15% to 20% more than that from CDL. There are exceptions like New Hampshire, Vermont and West Virginia which have lower proportion of land classified as cropland. Figure 4 shows a state-wise comparison of CDL and Agricultural Census data.

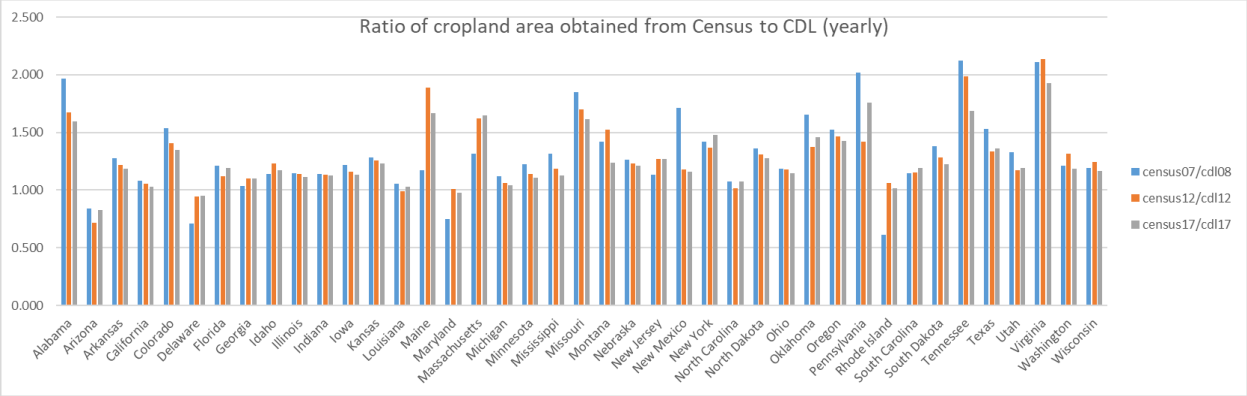

Figure 4: Comparison of cropland area – CDL and Agricultural Census

We also compare cropland estimations for USDA's Farm Resource Regions (FRR). There are 9 FRRs classified on the basis of agricultural intensity, types of crops produced, size of farms and other factors. Figure 5 shows a map of FRRs and table 3 compares cropland estimations from CDL and Agricultural Census for the FRRs.

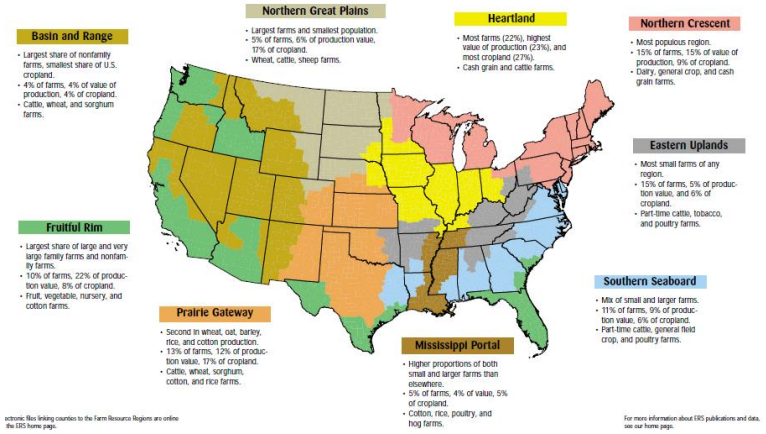

Figure 5: Farm Resource Regions (USDA)

As seen in table 3, CDL underestimates cropland area in the Appalachia. However, the differences are lower in the primarily agricultural areas like the Heartland and the Fruitful Rim.

Table 3: Ratio of cropland area from Agricultural Census to CDL for FRR's

FRR code	Farm Resource Region	Census07/CDL08	Census12/CDL12	Census17/CDL17
1	Heartland	1.2104	1.1754	1.1505
2	Northern Crescent	1.2776	1.2484	1.2260
3	Northern Great Plains	1.4398	1.3835	1.2979
4	Prairie Gateway	1.3568	1.2501	1.2598
5	Eastern Uplands	5.6154	3.8630	3.7196
6	Southern Seaboard	1.3172	1.3146	1.3104
7	Fruitful Rim	1.1421	1.1314	1.0931
8	Basin and Range	1.5591	1.5060	1.3650
9	Mississippi Portal	1.1029	1.0614	1.0253

In addition to gridcell level total cropland acreage and proportion, we also include in our analyses, acreage and proportion of major crops like corn, soybean, wheat and cotton among others. Crop parcels of same quality could have spatially different land rents due to accessibility of markets. In this paper we use Verburg et al. (2011)'s market accessibility index in order to control for such heterogeneity. The time invariant market accessibility index is generated by combining travel times to major cities, ports, and other towns and ranges from 0 to 1 with 1 being closest to the market. The extracted market accessibility index data are in 5 arcminute resolution and are aligned to our reference 5 arcminute gridcell raster. Figure 6 shows the map of market accessibility index with the bluer regions indicating higher indices. Table 4 gives summary statistics for the index and also the soil quality index. The market accessibility index is an approximation of how integrated and consequently urbanized a region is. However, in order to recognize gridcells that are completely within urban areas we use the 2019 Tiger/ Line Urban shapefiles from US Census. We find that out of the approximately 118,000 5 arcminute gridcells representing Contiguous US, 4113 gridcells can be classified as urban.

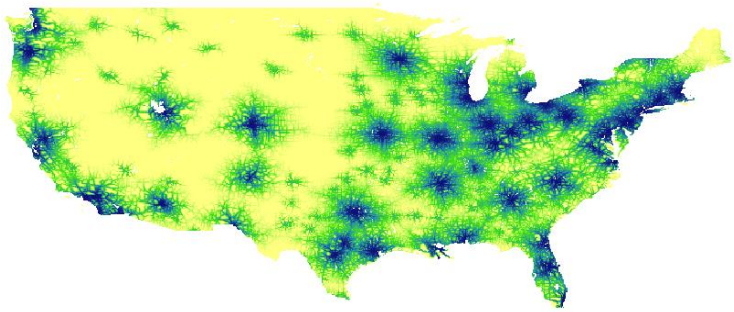

Figure 6: Market Accessibility Index – 0 (yellow) to 1(blue)

Cropland change with respect to cash rents depend also on the quality of soil. We use the National Commodity Crop Productivity Index (NCCPI) obtained from the USDA-NRCS Gridded Soil Survey Geographic (gSSURGO) Database product at 10m resolution. This index divides soil quality on the basis of numerous attributes like weather, soil carbon content, moisture, and crop type, and has a classification system with unique combinations of the attributes. These classifications are called MUKEY or map unit key and there are more than 306,000 soil classification categories for the US. The number of classification types (MUKEY) however is not equal to the number of unique values of the NCCPI soil index which ranges from 0 to 1. For example, while Kansas has 5793 MUKEY values, it has 723 unique values of the NCCPI soil

index. In order to extract the NCCPI data, we have reclassified the MUKEY to the unique NCCPI values using a Remap table in ArcGIS (which we created from the attribute table of the NCCPI raster). Since a raster value can only be reclassified to integer values, we next multiplied the NCCPI index values by 1000 (since it had upto three decimal places), aggregated them to 30 m cells, and finally used the raster calculator to divide the values by 1000. Next we use Raster Calculator to multiply each of the 30m NCCPI gridcells to that of the corresponding 30m CDL gridcells. Post multiplication, we reproject the raster to the same projection as the reference 5 arcminute gridcell raster (GCS WGS 1984). The reprojected raster is then aggregated using mean of each of the 30m cells and resampled to get 5 arcminute gridcell level cropland weighted NCCPI soil index raster. Figure 7 shows the original map of NCCPI in its native Albers Conical projection system to the left, and the reprojected 5 arcminute cropland weighted NCCPI to the right.

Figure 7: Soil Quality (NCCPI): 30m resolution (left) cropland weighted 5 arcminute NCCPI (right) (bluer regions indicate higher soil fertility)

For all the gridcells, the soil quality index ranges from 0 to 0.835 (most fertile) with a mean of 0.081 units (table 4).

Table 4: NCCPI and Market Accessibility Index

	#gridcells	Mean	Median	s.d.	Min	Max
NCCPI	119,753	0.081	0.004	0.149	0	0.835
Access Index	116,790	0.236	0.116	0.273	0	0.999

We obtain county specific cash rents from NASS-USDA. As mandated by the Farm Bill of 2008, NASS collects cash rental rates for counties that have at least 20,000 acres of land that is classified as cropland or pasture. Since there is a minimum acreage of cropland or pastureland required to be included in the survey of cash rents, several counties are not surveyed especially in 2008. Between 2009 and 2017, 70% of the counties had only non-irrigated cash rents, 25% had both irrigated and non-irrigated cash rents, and 4.5% had only irrigated cash rents available. We focus our study on the years 2009 through 2017 since cash rent surveys were not conducted in the years 2015 and 2018. A balanced panel dataset that includes either irrigated or non-irrigated cash rents along with other variables has 61,763 5 arcminute gridcells. Cash rents are at the county level. Thus, the gridcells belonging to a county have the same value for cash rents. Moreover, few counties and consequently gridcells have irrigated cash rents and thus estimating

separate land supply elasticities with respect to the change in the two types of cash rents would not capture the true variations in elasticities at the gridcell level. In order to obtain gridcell level estimates of cash rents, we utilize both irrigated and non-irrigated cash rents and combine them with gridcell level irrigation data.

We calculate gridcell level irrigation data from Moderate Resolution Imaging Spectroradiometer (MODIS)'s Irrigated Agriculture Datasets for the Conterminous United States (MIrAD-US) which is produced by USGS. The MIrAD-US data gives geospatial information on irrigated areas of the US. These geospatial datasets use three primary data inputs to produce and map the spatial distribution of irrigated land across contiguous US. The three primary data inputs are: (a) USDA's county-level irrigated area statistics for 2002, (b) annual peak eMODIS Normalized Difference Vegetation Index (NDVI), and (c) USGS derived land cover mask for agricultural lands from the National Land Cover Database (NLCD). The data are available every 5 years from 2002 to 2017, and we use the 250m resolution data in order to calculate the proportion of area irrigated for the 5 arcminute gridcells used in our study. The MIrAD-US datasets are first reprojected to the GCS-WGS-1984 projection system, then aggregated and resampled to 5 arcminute resolution, and finally merged to the gridcell level dataset. Since the irrigation data is available every 5 years, for the years 2008 through 2011, we assign irrigation data from 2007, for the years 2012 through 2016, irrigation data from 2012 is used, and for the year 2017 irrigation data from 2017 is used. Figure 8 shows the irrigated areas in the US in 2017. The Mississippi Alluvial Plains, the Central Valley, and parts of Central Great Plains and High Plains are some of the Eco-regions that are highly irrigated.

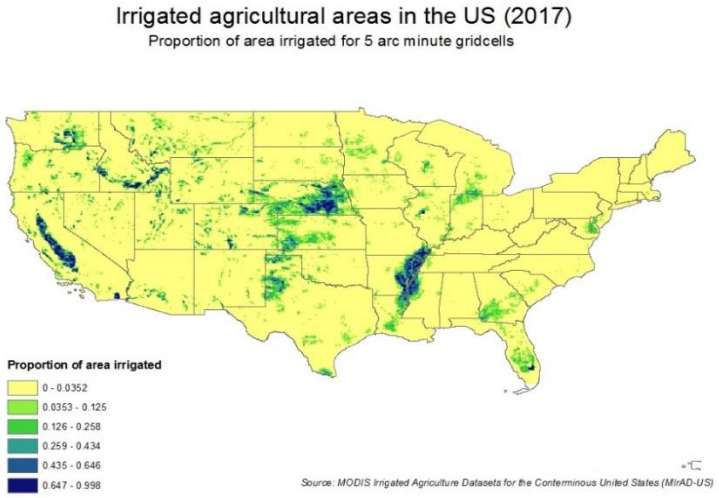

Figure 8: Irrigated areas of the US, 2017 (MIrAD-US)

We calculate gridcell level irrigation weighted cash rents by taking a mean of irrigated and non-irrigated cash rent weighted by proportion of area that is irrigated. The cash rents over the major agricultural Eco-regions Land rents increased between 2009 and 2013, especially in the Mid-West, but fell substantially between 2013 and 2017. Figure 9 shows the trends in irrigation weighted cash rents by primarily agricultural Eco-regions.

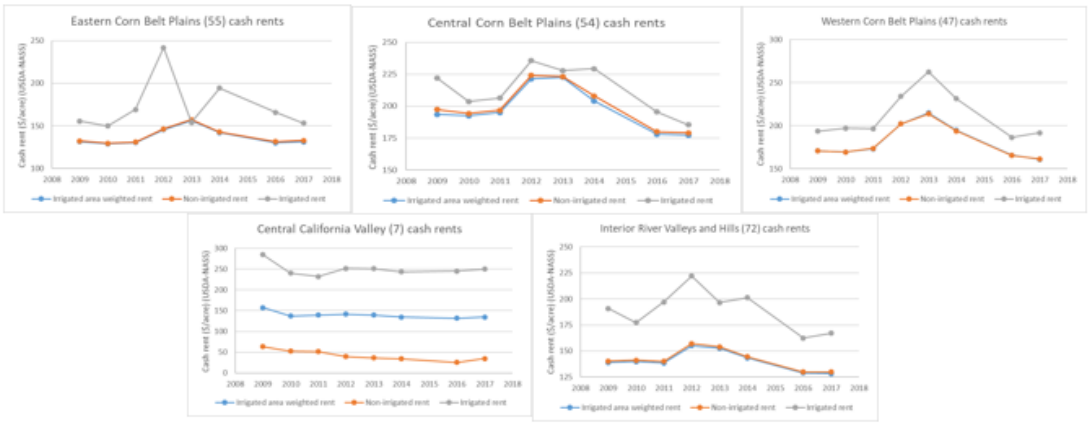

Figure 9: Cash rents by Eco-region

Decision to convert a parcel of land into cropland depends also on the amount of environmental benefits that a farmer could receive by conserving the land instead. One such program, the Conservation Reserve Program (CRP) is one of the largest voluntary land conservation programs in the US. The prevalent average CRP payments that a farmer could receive by enrolling their land into the program could determine how much land would be devoted to agriculture. We obtain county level average CRP payments per year from USDA's Farm Service Agency. Since there are several counties without information on CRP payments, the balanced panel dataset has 46,797 gridcells per year that are primarily located in the Mid-Western and other agricultural regions. Cash rents and average CRP payments per acre are at constant 2011 prices. Table 5.1 gives descriptive statistics of cash rents and CRP payments. The number of gridcells represents the total number of cells in the 8 year (2009-2017, excluding 2015) dataset. Although fewer gridcells have irrigated cash rents, on average irrigated cash rents are substantially larger than non-irrigated cash rents.

Table 5.1: Descriptive statistics of cash rents and CRP payments

	#gridcells	Mean	Median	s.d.	Min	Max
Irrigated cash rent	138,380	130.53	116.67	79.49	19.17	904.23
Non-irrigated cash rent	341,569	73.15	53.94	57.04	7.67	350.66
Irrigated area weighted cash rent	374,376	69.94	51.04	58.53	0	496.74
CRP average per acre	374,376	65.81	51.53	43.42	13.8	443.54
Rent to CRP ratio	374,376	1.032	1.035	0.64	0	19.53

For each county in our study, we also estimate the ratio of the cropland that is coming in or out of the CRP. We do it by first taking the negative of the net yearly change in land enrolled in the CRP program for each county. We call this delta CRP. For example, delta CRP for 2009 refers to the negative of the difference in CRP hectares in 2009 and 2008. We then calculate the net change in cropland in the county (calculated from the CDL) and calculate the change in Cropland as the difference in cropland hectares between two consecutive years (for example, 2009 and 2008). We call this delta cropland. When both the deltas move in the same direction, it implies that either CRP land is going into Cropland (both the deltas are positive), or cropland is going into CRP (both the deltas are negative). When the ratio is negative, this means that either cropland was expanding (Delta Cropland >0) and the land under CRP was also expanding (Delta CRP < 0) or vice versa. Table 5.2 shows the mean county level delta CRP and

cropland area, along with the mean of county level ratios of delta CRP to delta cropland. However, the means are not weighted by county area and therefore, the signs of the yearly change in mean county level CRP and cropland areas are not necessarily the same as those of the total CRP and cropland areas.

Table 5.2: Mean of yearly changes in CRP and cropland hectares and their ratio

Year	-ve Yearly change: CRP hectares (1)	Yearly change: cropland hectares (2)	Mean of county level ratio of 1 to 2
2008 - 2009	207.2	443.4	-1.1060
2009 - 2010	596.6	-2359.2	-0.0408
2010 - 2011	63.3	745.13	0.4576
2011 - 2012	451.7	1226.9	0.5070
2012 - 2013	692.4	-1444.0	0.3280
2013 - 2014	313.6	2381.4	-0.0003
2014 - 2016	117.1	460.4	0.3572
2016 - 2017	105.9	-216.69	5.1430

Impacts of climate change on crop yields is a well-documented phenomenon. With rising temperature extremes and average temperatures, the suitable temperature window for growing crops is getting reduced which in turn could decrease yields significantly in the long run (Schlenker and Roberts, 2009). The suitable temperature window conducive for growing crops vary by crops with the range being between 8°C and 32°C (Ritchie and NeSmith, 1991). Studies use the concept of degree days in order to capture how hot or cold a particular area is. A degree day compares mean temperatures at a given location to any standard temperature. In our study, we consider the suitable temperature window for crops to be between 10°C and 30°C and growing degree days (GDD) are calculated as "sum of truncated degrees between two bounds" as defined by Schlenker and Roberts (2009). We use Stata codes and PRISM Climate Group[3] datasets shared publicly by Wolfram Schlenker in his website[4] to estimate relevant degree days. The PRISM datasets give state-wise daily temperatures for 2.5 arcminute gridcells. While Schlenker's codes map these gridcells to counties and find average GDD for each county, we locate each of these 2.5 arcminute gridcells with reference to our 5 arcminute gridcells and find the average GDD for each of the 5 arcminute gridcells. In order to find the exact location of the 2.5 arcminute gidcells, we first create a buffer (circular) of 2.5 arcminute radius around each of the centroids of our reference 5 arcminute gridcells. Next, using the "Feature Envelope To Polygon" tool in ArcGIS, square polygons of 5 arcminute side for each of the centroids. Finally, we use the "Identity" tool to map the 2.5 arcminute PRISM gridcells to our 5 arcminute gridcells. There are 471,161 PRISM cells which are mapped to 121,280 5 arcminute reference gridcells. Figure 10 shows the process described above for a set of counties in Louisiana.

[3] http://www.prism.oregonstate.edu/
[4] http://www.wolfram-schlenker.com/dailyData.html

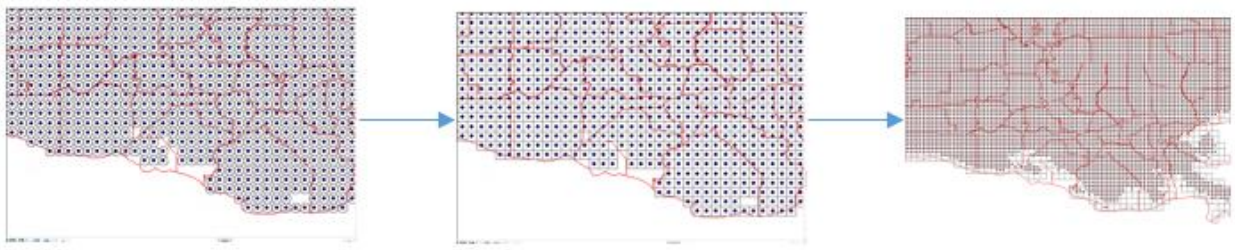

Figure 10: Mapping PRISM 2.5 arcminute cells to 5 arcminute reference gridcells
(Left and center: circular buffers and square polygons around centroids of 5 arcminute reference gridcells, Right: 2.5 arcminute PRISM cells mapped to 5 arcminute square polygons. Red lines indicate county boundaries)

Schlenker's codes also take into account the cropland proportion for each of the 2.5 arcminute PRISM cells. We assign uniform weights to all the PRISM cells since proportion of area under cropland is our dependent variable. The degree days are calculated for the temperature bounds 0°C, 10°C and 30°C. GDD, or total degree days between 10°C and 30°C, is calculated as follows using modified Schlenker's codes. If the average temperature for a day at a particular gridcell is 11°C, then the GDD for that day is 1 degree day. However, if the mean temperature is less than 10°C the degree days is 0. Thus, with 10°C as base, GDD can be a maximum of 20 degree days which corresponds to a mean temperature of 30°C. The degree days for the 2.5 arcminute PRISM cells are averaged by the 5 arcminute reference gridcells and the date for which each PRISM cell degree day is calculated for the temperature bounds. The degree days are then summed across 5 arcminute gridcells for each year. The GDD thus generated varies by gridcells and years. Very high temperatures have detrimental effects on yields (Asseng et al., 2015; Schlenker and Roberts, 2009). Therefore, the number of degree days for a threshold temperature that is above the suitable growing window of 10°C and 30°C, would indicate how much hot the temperature of a given location is such that it would negatively impact crop development. In this paper, we call this indicator the high-heat degree days (HHDD) which counts the number of degree days for mean temperatures above 30°C. The concept of HHDD is comparable to that of cooling degree days (CDD) used for calculating energy required for space cooling during higher temperatures. We also calculate gridcell level total yearly precipitation in millimeters using modified Schlenker's codes. Table 6 gives the mean (5 arcminute gridcell level) GDD, HHDD, and precipitation for the US for years 2009 through 2017. There is significant heterogeneity in all three indicators across states and counties. We also calculate climate normals, or averages of the preceding 30 years of climate data, for growing degree days, high heat degree days, precipitation, and average, minimum and maximum temperatures for all 5 arcminute gridcells.

Table 6: Mean GDD, HHDD and precipitation for the US

	2009	2010	2011	2012	2013	2014	2015	2016	2017
GDD (10<=c<=30)	1620.5	1720.8	1702.4	1789.7	1691.3	1660.8	1738.1	1743.0	1705.7
	642.7	692.2	688.3	623.5	597.8	630.4	627.8	637.9	620.0
HHDD (>30c)	45.8	56.2	76.5	68.9	48.0	41.6	49.0	54.1	48.6
	84.1	78.8	108.1	85.8	78.2	74.3	76.1	79.5	81.4
Precipitation (mm/year)	452.1	466.6	457.8	374.4	472.5	457.8	473.5	453.4	471.4
	269.9	239.4	277.5	238.6	263.2	236.1	252.4	256.5	293.5

There is evidence that cropland (especially corn) area has increased in response to bio-fuel boom during the late 2000's (Motamed et al., 2016). Moreover, location of bio-fuel plants had a higher impact on crop expansion in regions where previous cropland areas were low. Thus, location of bio-fuel plants could be a

determinant of land supply elasticities. In this paper, we use data on biofuels plants, specifically ethanol and bio-diesel plants, and use the location, vis-à-vis the 5 arcminute reference gridcells, and operational status of the plants as independent variables. We obtain the location and operational status data from the National Bio-refineries Database[5] which has cross-sectional data for the year 2009. We verify the exact coordinates of the bio-fuel plants using their addresses and combine the various operational status types into 5 categories – Construction, Idle, Operational, Shutdown/bankrupt, and Unknown. Table 7 gives the distribution of bio-fuel plants by operational status. There were more number of bio-diesel plants in the US in 2009. However, the proportion of operational plants was lower than that of ethanol plants.

Table 7: Bio-fuel plants by operational status

Status	Bio-diesel	Ethanol
Construction	30	8
Idle	10	9
Operational	174	144
Shutdown/bankrupt	12	4
Unknown	113	32
Total	339	197

Figure 11 shows the distribution of ethanol and bio-diesel plants across the US irrespective of operational status. While most the ethanol plants are concentrated in the heartland region, the distribution of bio-diesel plants is more uniform.

Figure 11: Left: Ethanol plants; Right: Bio-diesel plants

Preliminary results

We estimate land supply elasticities from the panel specification in equation 3 and using the formula in equation 4. Additionally, we also estimate elasticities using a long difference estimator. An estimator for two time periods separated by 4 years takes the following form:

$$\log Z_{it+4} - \log Z_{it} = \beta_0 + \beta_1(\log R_{it+4} - \log R_{it}) \qquad (5)$$

The estimated coefficient $\hat{\beta}_1$ is the land supply elasticity or percentage change in cropland proportion for a 1% increase in land rent between the years *t* and *t+4*. Z_{it} refers to cropland proportion, and R_{it} refers to irrigation weighted cash rent for gridcell *i* in time *t.* We compare the years 2009 to 2013, 2013 to 2017, and 2009 to 2017. Proximity of gridcells to one another could have impacts on land supply elasticities. Therefore, in order to avoid spatial autocorrelation, we take a stratified (by Eco-region) 25% sample of gridcells, and estimate land supply elasticities. Table 8 summarizes US and Farm Resource Region level

[5] https://openei.org/datasets/dataset/national-biorefineries-database

elasticities estimated using both panel and long difference estimations. For panel estimations, for the period 2009-17, the land supply elasticity was 0.0107. While there is heterogeneity in results in terms of both Farm Resource Region and time periods used in estimation, out results are preliminary and the specifications use access index, soil quality and prevalent non-cropland land use type as the only independent variables in addition to irrigation weighted cash rents. The results are expected to change significantly when we include time variant gridcell level variables like growing degree days and precipitation among others.

Table 8: Preliminary results using a stratified (by Eco-region) 25% sample of gridcells

FRR Code	Farm Resource Region	Panel Estimation Elasticities						Long Difference Elasticities					
		2009-13	n	2013-17	n	2009-17	n	2009-13	2013-17	2009-17	n	2 period	n
1	Heartland	-0.228***	11,835	0.184***	9,468	-0.00717	18,936	-0.0322	-0.0265	-0.240***	2,620	0.0278	5,240
2	Northern Crescent	-0.0416***	7,885	0.317***	6,308	0.0449***	12,616	0.0958***	0.0147	0.291***	1,845	0.0269**	3,690
3	Northern Great Plains	0.0467***	8,740	0.0406*	6,992	0.0533***	13,984	0.0670***	-0.0318	0.0253	2,227	0.0214	4,454
4	Prairie Gateway	0.0627***	9,920	-0.0686***	7,936	0.0133***	15,872	0.0393*	-0.0591***	0.0275	2,455	-0.0165	4,910
5	Eastern Uplands	0.170***	3,515	-0.0323	2,812	0.212***	5,624	0.450***	-0.528***	0.235**	1,053	-0.0254	2,106
6	Southern Seaboard	-0.133***	5,455	0.246***	4,364	-0.014	8,728	0.200***	0.0953***	0.115**	1,324	0.174***	2,648
7	Fruitful Rim	-0.131***	5,670	0.197***	4,536	-0.00735	9,072	0.0224	-0.036	0.0371	1,768	-0.0138	3,536
8	Basin and Range	-0.0812*	5,195	0.0983	4,156	0.00237	8,312	-0.014	-0.0116	0.0155	1,472	-0.0183	2,944
9	Mississippi Portal	0.0693	2,655	-0.103**	2,124	-0.0112	4,248	0.0954	0.0165	0.0114	681	0.0649	1,362
	USA	-0.0643***	60,870	0.0752***	48,696	0.0107***	97,392	0.0458***	-0.026***	0.0619***	15,445	0.00127	30,890

Figure 12 shows maps for land supply elasticities at the level of eco-regions using full sample long differences estimator. Significant heterogeneities are observed by eco-region as well as time periods. While the regions around the Northern Great Plains have consistently high land supply elasticites, eco-regions in the Hearland region like the Western Corn Belt in Iowa have inconsistent elasticities. We also find that the traditionally highly irrigated eco-regions like the Mississippi Alluvial Plain, the Central Valley in California, and the Snake River Plain in Idaho have low land supply elasticities.

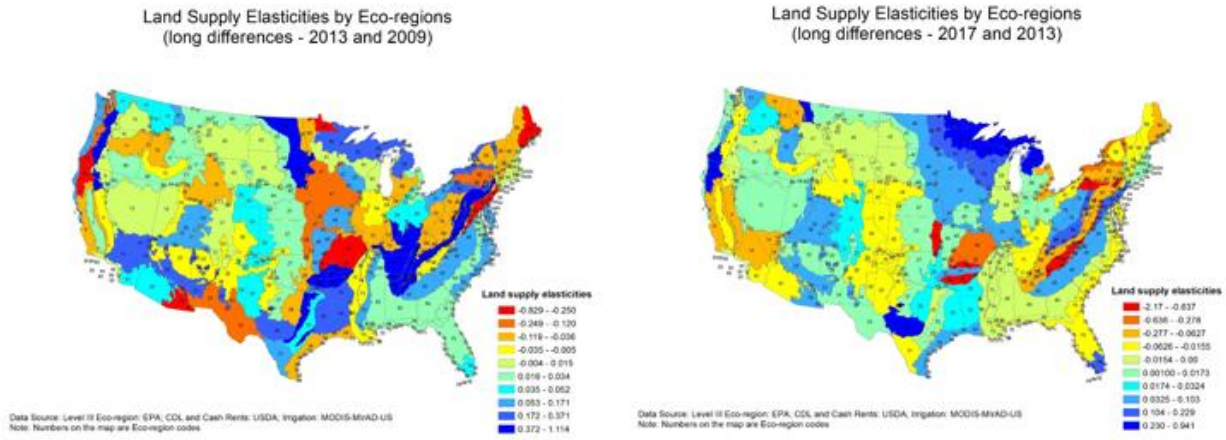

Figure 12: Land supply elasticities by eco-regions estimated using long differences estimator

One of the primary objectives of our study is to check the predictability power of the estimated land supply elasticities. In order to test the predictive power of our land supply elasticities, we compute correlations between predicted and original proportion of area under cropland. Estimates from the long difference

models using 25% sample are used to predict cropland proportion of another 25% sample of the data. Although both Spearman and Pearson correlation coefficients are less than 0.3, they are significant and out-of-sample predictive power are relatively higher for the 2009-13 and 2009-17 models (table 9). For the Northern Crescent and the Northern Great Plains, the predictability power is relatively higher. In terms of Eco-regions, predictive power is higher for Great Plains and Glaciated Plains, but lower for Western and Central Corn Belt Plains.

Table 9: Out of sample predictions (using 25% stratified by Eco-region sample)

		Spearman rank correlation coefficient				Pearson correlation coefficient			
FRR Code	Farm Resource Region	2009-13	2013-17	2009-17	n	2009-13	2013-17	2009-17	n
1	Heartland	0.0502*	-0.2123*	-0.0445*	2588	0.1216*	0.0024*	0.19*	2,588
2	Northern Crescent	0.1429*	0.1305*	0.1695*	1829	0.0921*	0.0371*	0.1623*	1,829
3	Northern Great Plains	0.2655*	0.1339*	0.2175*	2285	0.0418*	0.0531*	0.0726*	2,285
4	Prairie Gateway	0.1733*	0.033*	0.1543*	2405	0.0233*	0.0158*	0.0783*	2,405
5	Eastern Uplands	0.1452*	0.1349*	0.1289*	1073	0.1004*	0.1444*	0.0182*	1,073
6	Southern Seaboard	0.1575*	0.0442*	0.1322*	1343	0.0741*	0.0494*	0.083*	1,343
7	Fruitful Rim	-0.0099	0.0303*	0.0296*	1765	0.0235*	0.0808*	0.0124	1,765
8	Basin and Range	0.007	0.0643*	0.0337*	1505	0.019*	0.0182*	0.0081	1,505
9	Mississippi Portal	0.1876*	0.0315*	0.107*	647	0.1385*	0.0242*	0.0537	647
	USA	0.2701*	0.1509*	0.2385*	15,440	0.2268*	0.1171*	0.2051*	15,440

We plan to test the predictive power of the remaining models (including both long difference and panel models), and also incorporate additional time varying covariates like precipitation and temperature that could impact cropland change.

References

Asseng, S., Ewert, F., Martre, P., Rötter, R. P., Lobell, D. B., Cammarano, D., ... & Reynolds, M. P. (2015). Rising temperatures reduce global wheat production. *Nature climate change*, *5*(2), 143.

Baldos, U. L. C., & Hertel, T. W. (2013). Looking back to move forward on model validation: insights from a global model of agricultural land use. Environmental Research Letters, 8(3), 034024.

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

Goolsby, D. A., Battaglin, W. A., Aulenbach, B. T., & Hooper, R. P. (2001). Nitrogen input to the Gulf of Mexico. *Journal of Environmental Quality*, *30*(2), 329-336.

Graesser, J., Aide, T. M., Grau, H. R., & Ramankutty, N. (2015). Cropland/pastureland dynamics and the slowdown of deforestation in Latin America. *Environmental Research Letters*, 10(3), 034017.

Hellerstein, D., & Malcolm, S. (2011). The influence of rising commodity prices on the Conservation Reserve Program. Economic Research Service, Paper No. ERR, 110.

Lark, T. J., Mueller, R. M., Johnson, D. M., & Gibbs, H. K. (2017). Measuring land-use and land-cover change using the US department of agriculture's cropland data layer: Cautions and recommendations. *International journal of applied earth observation and geoinformation*, *62*, 224-235.

Lark, T. J., Salmon, J. M., & Gibbs, H. K. (2015). Cropland expansion outpaces agricultural and biofuel policies in the United States. *Environmental Research Letters*, 10(4), 044003.

Motamed, M., McPhail, L., & Williams, R. (2016). Corn area response to local ethanol markets in the United States: A grid cell level analysis. *American Journal of Agricultural Economics*, *98*(3), 726-743.

Papke, L. E., & Wooldridge, J. M. (2008). Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics*, 145(1-2), 121-133.

Ritchie, J. T., & Nesmith, D. S. (1991). Temperature and crop development. *Modeling plant and soil systems*, *31*, 5-29.

Schlenker, W., & Roberts, M. J. (2009). Nonlinear temperature effects indicate severe damages to US crop yields under climate change. *Proceedings of the National Academy of sciences*, *106*(37), 15594-15598.

Turner, R. E., & Rabalais, N. N. (1991). Changes in Mississippi River water quality this century. *BioScience*, 41(3), 140-147.

Verburg, P. H., Ellis, E. C., & Letourneau, A. (2011). A global assessment of market accessibility and market influence for global environmental change studies. *Environmental Research Letters*, *6*(3), 034019.

Villoria, N.B., Liu, J., 2018. Using continental grids to improve understanding of global land supply responses: Implications for policy-driven land use changes in the Americas. Land Use Policy 75, 411–419

Zhu, J., Morgan, G.D., 2004. Comparison of spatial variables over subregions using a block bootstrap. *J. Agric. Biol. Environ. Stat.* 9, 91–104