

# Econometric Estimates of Land Supply Functions Using Continental Gridded Data



Nelson Villoria

Center for Global Trade Analysis  
Department of Agricultural Economics  
Purdue University

June 23, 2012

# Motivation

# Research Gaps

- Treatment of the extensive margin in CGE models is constrained by lack of econometric evidence on land supply elasticities as well as by functional forms
- Econometric studies (whether focusing on yield response, land rents, or crop choices) assume land as fixed. [e.g., Cline (2007); Mendelsohn and Seo (2007); Lobell, Schlenker, and Costa-Roberts, 2011; Schlenker and Roberts, 2006].
- To my knowledge, only two studies attempt to estimate country level land supply elasticities: Lubowsky (2002) for the US and Barr et al (2011) for the US and Brazil.
- Only Lubowsky effectively deals with land heterogeneity.

# Research Objective

Exploit global gridded datasets on land cover and agricultural productivity to estimate land supply curves that take into account land quality heterogeneity

Theory

# The Spatially-Explicit Von Thünen Model of Land Use

Assume there are two land uses: agriculture and natural state.

Let  $i$  ( $i = 1, \dots, N$ ) index the  $i^{th}$  grid cell.

The farmer's problem managing grid cell  $i$  is to:

$$\max_{l_i} Y(l_i, \mathbf{H}_i, \mathbf{M}_i)P_i - l_i V_i$$

where  $Y_i = Y(\cdot)$  is the per-area production function. The optimal land demand,  $l_i^*(P_i, V_i, \mathbf{H}_i, \mathbf{M}_i)$  gives the (short-run or restricted) profit function as a function of input and output prices:

$$\pi_i(P_i, V_i, \mathbf{H}_i, \mathbf{M}_i)$$

. Above and in what follows:

- $Y$  are yields
- $l$  is a non-land input composite
- $\mathbf{H}$  is a vector of fixed factors
- $\mathbf{M}$  is a vector of managerial factors and other socioeconomic shifters

# Unobservable Prices and Von Thünen's Land Rents

We do not observe prices. Von Thünen's key assumption is that spatial differentials in farm-gate prices are related solely to differences in transport costs to major markets:

- Output price received at grid cell  $i$  is  $P_i = P(P, F_i, D_i)$
- Input price paid is  $V_i = V(V, F_i, D_i)$
- $\frac{\partial P_i}{\partial D_i} < 0$  and  $\frac{\partial V_i}{\partial D_i} > 0$ , where:
  - $P$  is the price of the output composite
  - $V$  is the price of the non-land input composite
  - $F$  are transportation costs
  - $D$  is distance to markets

For our purposes, land is the only fixed factor of production, and thus

$$\pi_i(.) \equiv R_i^A(F_i, D_i, \mathbf{H}_i, \mathbf{M}_i) = R_i^A(\mathbf{X}_i)$$

are the so-called von Thünen rents.

# Implications of the Von Thünen Model

- Both input intensity and productivity decrease with distance (which implies that the effect of  $D$  on  $Y$  is a testable hypothesis)
- Land will be employed in the activity yielding the highest rent. In other words, land at grid cell  $i$  will be devoted to productive use  $A$  if and only if:

$$R_i^A > R_i^N$$

Assuming the farmer knows  $R_i^A$  and  $R_i^N$  her problem becomes:

$$\max_{Z_i} Z_i R_i^A + (1 - Z_i) R_i^N$$

where the optimal discrete choice of allocating land to agriculture or leaving it in its natural use is given by:

$$Z_i^* = Z(\mathbf{X}_i).$$

# Econometrics

# Econometric Implementation: Three Main Issues

- 1 Spatial autocorrelation:
  - relevant variables tend to be similar across neighboring units  
→ shocks in one location are likely correlated with shocks in nearby locations
- 2 Omitted variable bias:
  - sparse data environment where many key determinants of land use, e.g., prices and policies, are unobserved
- 3 Land accessibility and land use are likely to be determined simultaneously, at least in some locations

Issues 1 and 2 are partially tackled by explicit modeling of spatial relationships in the data; potential for access/land-use endogeneity is ignored.

# Spatial Autocorrelation in Discrete Choice Models

## Difficulties:

- Standard spatial models typically imply heteroscedasticity and autocorrelation rendering the ML estimator inconsistent
- Alternative consistent estimators have been proposed, but they are infeasible with large datasets because they require repeated inversion of  $n \times n$  matrices, where  $n$  is the sample size
- I use a recent GMM estimator proposed by Klier and McMillen (2008, JBES, 26:4, pp: 460-471):
  - Estimation does not require matrix inversion so can be used in large samples
  - However, obtaining marginal effects requires inversion of a  $n \times n$  matrix once effectively reducing feasible sample sizes

# Spatial Logit Model

We assume that the probability  $Z_i^* = 1$  follows the logistic distribution  $\Lambda$ , so, the econometric model is:

$$P(\mathbf{Z} = 1 | \mathbf{X}, \mathbf{W}\mathbf{y}^*, \mathbf{W}\mathbf{X}^L) = \Lambda \left( \rho \mathbf{W}\mathbf{y}^* + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}^L\boldsymbol{\theta} \right) \quad (2)$$

where:

- $\mathbf{Z}$  and  $\mathbf{y}^*$  are the  $n \times 1$  vector of land use decisions and unobservable returns to agriculture discussed above
- $\mathbf{W}$  is a  $n \times n$  row-standardized (e.g.,  $\sum_j w_{ij} = 1$ ) weight matrix defining neighborhood relationships between grid-cells
- $\mathbf{X}$  is a  $n \times k$  matrix of regressors
- $\mathbf{X}^L$  is a  $n \times l$  matrix where  $l$  is a subset of the  $k$  regressors in  $\mathbf{X}$ .
- $\mathbf{W}_i\mathbf{y}^*$  is the weighted average of the unobservable land returns of gridcell  $i$ 's neighbors
- $\rho$  is the so-called auto-regressive spatial parameter
- $\mathbf{W}_i\mathbf{X}^L$  captures the weighted average of the explanatory variables in the neighborhood of gridcell  $i$
- $\boldsymbol{\theta}$  is as  $l \times 1$  is a the parameter vector associated with the lagged explanatory variables
- $\boldsymbol{\beta}$  is the  $k \times 1$  vector of parameters summarizing the effect of  $\mathbf{X}$  on  $\mathbf{y}^*$ .

# Marginal Effects and Elasticities

The marginal effect of market access on the probability of land use is given by:

$$\frac{\partial P(Z_i = 1|.)}{\partial ma_i} = \lambda(\rho \mathbf{W}\mathbf{y}^* + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}^L\boldsymbol{\theta}) \left( \hat{\beta}_{ma} \right) \quad (3)$$

where  $\lambda$  is the probability density function of the logistic distribution. The elasticity to market access is given by:

$$\epsilon_{ma_i} = \lambda(.) \left( \hat{\beta}_{ma} \right) \frac{ma_i}{P(Z_i = 1|.)}. \quad (4)$$

As the equations above evidence, the Logit model produces marginal effects and an associated elasticities for each observation. With more than 15,000 gridcells used for estimation this is simply unmanageable. Thus, we aggregate the results using meaningful observation units such as countries and agroecological zones.

Data

# Regressand

- Based on harvested area (as % of each gridcell) from Monfreda, Ramankutty and Foley (2008).
- Circa year 2000
- 5 min ( $\approx 9.16$  km/5.69 miles) by 5 min latitude-longitude grid covering the world.
- The dependent variable is an indicator variable:
  - $Z=1$  if more than 5% of gridcell is under crops
  - $Z=0$  otherwise

# Regressors

| VARIABLES      |                                 | UNITS          | SOURCE             |
|----------------|---------------------------------|----------------|--------------------|
| Biophysical    | Potential Vegetation*           | 1-14           | Ramankutty & Foley |
|                | Soil Fertility Constraints      | 1-7            | IIASA Global AEZ   |
|                | Average annual precipitation    | mm             | IIASA Global AEZ   |
|                | Elevation                       | meters         | NOAA SAGE's Atlas  |
|                | Soil pH                         | 0-14           | SAGE               |
|                | Soil Carbon                     |                | SAGE               |
|                | Monthly temp (ave. 1961-1990)** | Degree celsius | CRU                |
| Socio-economic | Market Access                   | index          | Verburg et al      |
|                | Area equipped for irrigation    | % of gridcell  | Siebert et al      |
|                | Built-up land                   | % of gridcell  | Miteva, B.         |
|                | Protected Areas                 | 0-2            | GEONETWORK         |
| Descriptive    | GTAP 18 and 108 AEZs            | AEZs           | Uris Baldos        |

\* Reduced to five categories: Temperate Forests, Tropical Forests, Grasslands and Savanna, Shrublands, Other (Tundra, Desert, etc.)

\*\* Based on monthly temperatures 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 |

# Sampling

- We focus on the Americas because it provides a geographic continuum across the North-South axis with enough variability across biophysical variables.
- Sample limited by latitudes below 55.625N.
- Islands are discarded, 23 countries including: Canada, United States, Mexico, Guatemala, Honduras, Nicaragua, El Salvador, Colombia, Venezuela, Costa Rica, Panama, Guyana, Suriname, French Guiana, Brazil, Ecuador, Peru, Bolivia, Chile, Paraguay, Argentina, and Uruguay.
- Remaining sample size,  $n=433,096$ , is still too large for appropriately dealing with spatial autocorrelation.
- Sample (regularly) one gridcell every 30 minutes, for a sample size of approx. 15,000.

# Some Descriptive Statistics: Sample Representativeness

Proportion of gridcells used in agriculture in each land cover type (population vs. sample)

|                        | Population | Sample |
|------------------------|------------|--------|
| Other (Tundra, Desert) | 0.03       | 0.03   |
| Forests                | 0.23       | 0.22   |
| Grassland/Savanna      | 0.53       | 0.52   |
| Shrublands             | 0.27       | 0.26   |

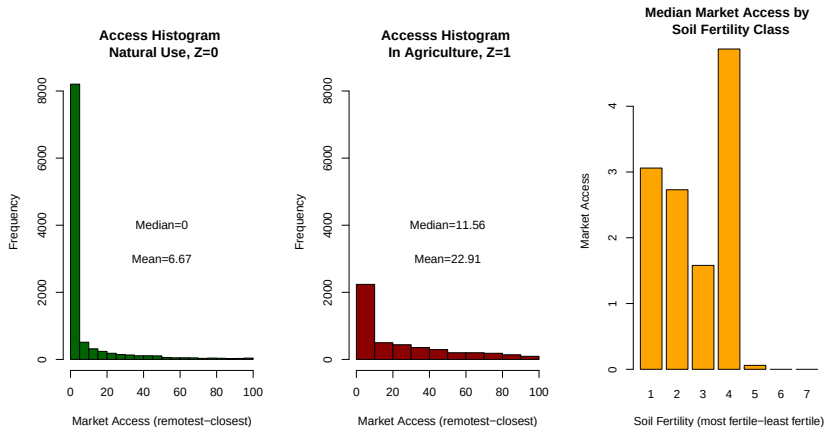
Proportion of gridcells used in agriculture in each soil fertility category (1 is most fertile, population vs. sample)

|           | Population | Sample |
|-----------|------------|--------|
| 1 (most)  | 0.58       | 0.58   |
| 2         | 0.40       | 0.40   |
| 3         | 0.36       | 0.35   |
| 4         | 0.41       | 0.40   |
| 5         | 0.24       | 0.24   |
| 6         | 0.12       | 0.13   |
| 7 (least) | 0.14       | 0.12   |

Proportion of gridcells under agricultural use in each AEZ (population vs. sample)

|    | Population | Sample |
|----|------------|--------|
| 1  | 0.19       | 0.18   |
| 2  | 0.31       | 0.38   |
| 3  | 0.43       | 0.41   |
| 4  | 0.37       | 0.37   |
| 5  | 0.23       | 0.23   |
| 6  | 0.13       | 0.12   |
| 7  | 0.25       | 0.24   |
| 8  | 0.54       | 0.52   |
| 9  | 0.38       | 0.38   |
| 10 | 0.63       | 0.63   |
| 11 | 0.73       | 0.74   |
| 12 | 0.57       | 0.56   |
| 13 | 0.20       | 0.18   |
| 14 | 0.05       | 0.06   |
| 15 | 0.08       | 0.08   |
| 16 | 0.31       | 0.35   |

# A closer look at the market access index



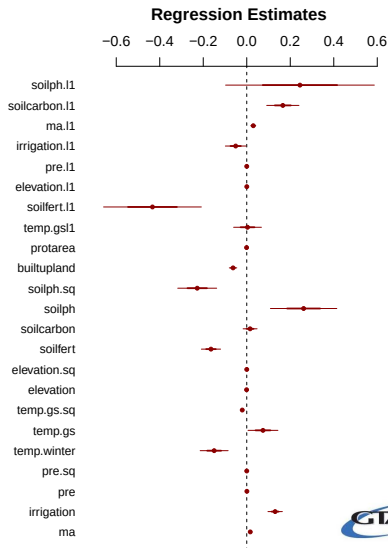
Market access distribution is highly skewed, with most gridcells being far away from markets. Generally speaking, gridcells already in agriculture tend to be closer to markets than gridcells still in their natural use. The plot in the right suggests that, the least fertile gridcells (class 5-7) are very far away from markets. To the extent that the most fertile gridcells tend to have a higher use incidence, this suggests that market access and land use are to some extent simultaneously determined. Although we keep in mind, we have little room for dealing with this endogeneity issue in the present data environment/study design.

# Results

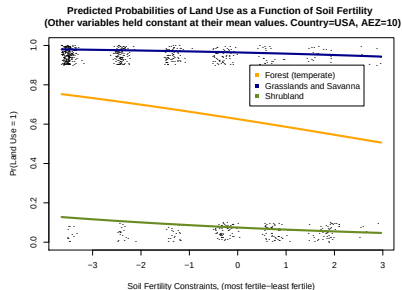
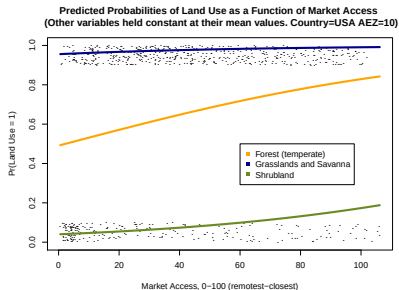
# Spatial Logit Regression Results

- In the plot, dots are point estimates and the segments are 90% and 95% confidence intervals
- 15,093 observations
- Cover, country, and AEZ specific intercepts included but not shown.
- Spatial autoregressive parameter: 0.1243, non-significant
- Model rightly predicts  $Z=0$  in 80% of the 10,459 observed cases and it rightly predicts  $Z=1$  in 83% of the 4,634 observed cases.

Market access ("ma") has a positive, significant, albeit moderate, effect on the probability of land being used in agriculture. Most of the included variables are significant and have expected signs. Variables with suffix ".l1" are spatially weighted averages of neighboring values.

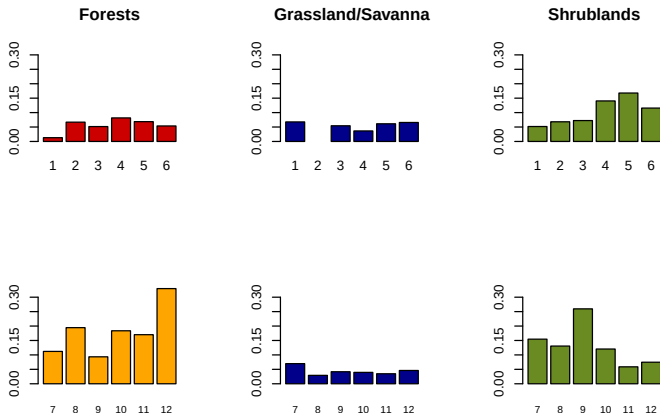


# Regression Results: Accessibility vs. Land Quality



The model can be used to identify (i) the effects of market access on the responsiveness of each cover type while holding land quality and other variables constant (left), or (ii) the effects on land quality on the probability of land use holding market access and other variables constant (right).

# Land Supply Elasticities by Land Cover and AEZ



AEZs 1-6 are tropical and 7-12 are temperate. The reported values are AEZ-cover averages weighted by the predicted probability of land use. The only work we have to put these estimates in perspective is Lubowski (2002), who focuses on the US. His forest-to-cropland elasticity (from a nested Logit specification, evaluated at the means) is 0.2104, our analogous estimate is 0.2339. His pasture-to-cropland elasticity (conditional Logit, evaluated at the means) is 0.0598. Ours is 0.0572.

# Concluding Remarks

# Concluding Remarks

- We set out to estimate land supply elasticities using large-scale gridded data and succeeded capturing differences across land cover types, AEZs, and countries.
- To the extent that they are comparable, my estimates are in line with previous literature.
- Estimation of the model proved to be laborious and replication in other geographic areas is likely to require as much effort as the work presented here. My following task is to explore how good the fitted model predicts land use in other environments.

# For Discussion

- Now that we have parameters by AEZ and cover, can we use them to improve the parametrization of conversion across land types? Could this extra information on elasticities help moving GTAP-AEZ beyond the CET?
- More importantly, can we use these elasticities to bring lands under natural uses into production thus increasing land endowment?
- That is, can we infer from price increases (in this framework equated to access improvements) how much more likely are unused lands to transition into agriculture?
- Once a parcel transition, could not we assign a potential rent based on latent productivity?

# References



Mendelsohn, R., W. D. Nordhaus, and D. Shaw. 1994. "The Impact of Global Warming on Agriculture: A Ricardian Analysis." *The American Economic Review* 84, no. 4: 753-771.



Mendelsohn, R. O., and N. Seo. 2007. "Changing Farm Types and Irrigation as an Adaptation to Climate Change in Latin American Agriculture." World Bank Policy Research Working Paper No. 4161, SSRN. Washington, DC.



Cline, W. R. 2007. *Global Warming and Agriculture: Impact Estimates by Country*. Washington DC: Center for Global Development.



Lee, H.-L., T. W. Hertel, B. Sohngen, and N. Ramankutty. 2005. "Towards an Integrated Land Use Data Base for Assessing the Potential for Greenhouse Gas Mitigation." GTAP Technical Paper No. 25, Purdue University. West Lafayette.



Schlenker, W., and M. J. Roberts. 2006a. Estimating the Impact of Climate Change on Crop Yields: The Importance of Non-Linear Temperature Effects: SSRN.



Schlenker, W., and M. J. Roberts. 2006b. "Nonlinear Effects of Weather on Corn Yields." *Review of Agricultural Economics* 28, no. 3: 391-398.



Klier, Thomas, and Daniel P McMillen. "Clustering of Auto Supplier Plants in the United States." *Journal of Business & Economic Statistics* 26, no. 4 (October 2008): 460-471.



Lubowski, Ruben (2002) Determinants of Land Use Transitions in the United States: Econometrics Analysis of Changes among the Major Land-Use Categories. PhD Dissertation, Harvard University: Cambridge, MA.



Barr, Kanlaya J., Bruce A. Babcock, Miguel A. Carriquiry, Andre M. Nassar, and Leila Harfuch. Agricultural Land Elasticities in the United States and Brazil. *Applied Economic Perspectives and Policy* 33, no. 3 (Autumn 2011): 449-462.