Reconciling econometric and simulation models of agricultural supply using a Generalized Extreme Value model

Christophe Gouel[1,2] David Laborde[3] Simone Pieralli[1]

2019 GTAP Conference

[1]INRA, Économie Publique

[2]CREST, Ecole Polytechnique

[3]IFPRI, Market, Trade, and Institutions

Introduction

Motivation

- Land-use related questions now important in CGE work
- But their land-use specifications (based on nested CET functions) have evolved little
- 3 main problems:
  - Do not respect physical balances
    - $\Rightarrow$ No real concept of yields
  - Ad hoc spatial disaggregation
  - Nesting and calibration of elasticities based on limited evidence

Motivation

- Land-use related questions now important in CGE work
- But their land-use specifications (based on nested CET functions) have evolved little
- 3 main problems:
  - Do not respect physical balances
    - ⇒ No real concept of yields
  - Ad hoc spatial disaggregation
  - Nesting and calibration of elasticities based on limited evidence

Contributions

This paper shows

- How to solve these problems using a modeling based on Generalized Extreme Value (GEV) distributions
- While being very close to the existing modeling based on nested CET functions
- How to make the connection with the other existing approaches in the AgEcon literature: PMP and multi-crop econometric models.

Simple model

Setup

- Land endowment $\bar{L}$
- Crops indexed by $k \in \mathcal{K} \equiv \{1, \ldots, K\}$
- Crops only require land to grow.

CET modeling i

- Production: $Q_k = A_k L_k$
- CET function with $\eta < 0$:

$$\bar{L} = \left(\sum_{k \in \mathcal{K}} \gamma_k L_k^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}},$$

- The farmer maximizes his profits $\sum_{k \in \mathcal{K}} p_k A_k L_k = \sum_{k \in \mathcal{K}} R_k L_k$ subject to $\{L_k\}_{k \in \mathcal{K}}$

CET modeling ii

- FOC:

$$L_k = \gamma_k^{\eta} \left(\frac{R}{R_k} \right)^{\eta} \bar{L},$$

$$R = \left(\sum_{k \in \mathcal{K}} \gamma_k^{\eta} R_k^{1-\eta} \right)^{\frac{1}{1-\eta}}.$$

- In consequence:

$$\pi_k = \frac{R_k L_k}{R \bar{L}} = \frac{\gamma_k^{\eta} R_k^{1-\eta}}{\sum_{l \in \mathcal{K}} \gamma_l^{\eta} R_l^{1-\eta}},$$

$$Q_k = \bar{L} A_k \gamma_k^{\eta/(1-\eta)} \pi_k^{-\eta/(1-\eta)}.$$

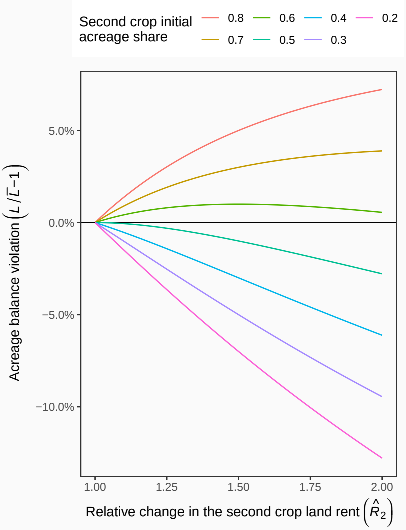
Lack of respect of physical balances with the CET

1^{st}-order approximation of physical balance

With

- $L = \sum_{k \in \mathcal{K}} L_k$: physical endowments
- π_k: land-revenue share
- ϖ_k: land-use share

$$\mathrm{d} \ln \frac{L}{\bar{L}} = \eta \sum_{k \in \mathcal{K}} (\pi_k - \varpi_k) \, \mathrm{d} \ln R_k.$$



Fréchet-based modeling i

- Heterogeneous land composed of a continuum of parcels indexed by $\omega \in [0,1]$.
- Production per unit of land on parcel ω if planted crop k: $Q_k(\omega) = A_k(\omega)$
- Yields, $A_k(\omega)$, are i.i.d. from a Fréchet distribution with shape $\theta > 1$ and scale $\gamma A_k > 0$:

$$\Pr(A_k(\omega) \leq a) = \exp\left[-\left(\frac{a}{\gamma A_k}\right)^{-\theta}\right] \quad \forall a \in \mathbb{R}_{>0}.$$

such that $A_k = \mathrm{E}[A_k(\omega)]$.

Fréchet-based modeling ii

- Producer pb on parcel ω: plant the most profitable crop (i.e., highest land rents $R_k(\omega) \equiv p_k A_k(\omega)$) $\Leftrightarrow$ discrete choice pb
- Probability that k is the most profitable defined by

$$\pi_k = \Pr\left(R_k(\omega) \in \arg\max_{l \in \mathcal{K}} R_l(\omega)\right), \\ = \frac{R_k^\theta}{\sum_{l \in \mathcal{K}} R_l^\theta} = \left(\frac{R_k}{R}\right)^\theta,$$

where $R_k \equiv p_k A_k$ is the unconditional land rent.

- Since identical parcels, $\pi_k =$ share of land allocated to k.

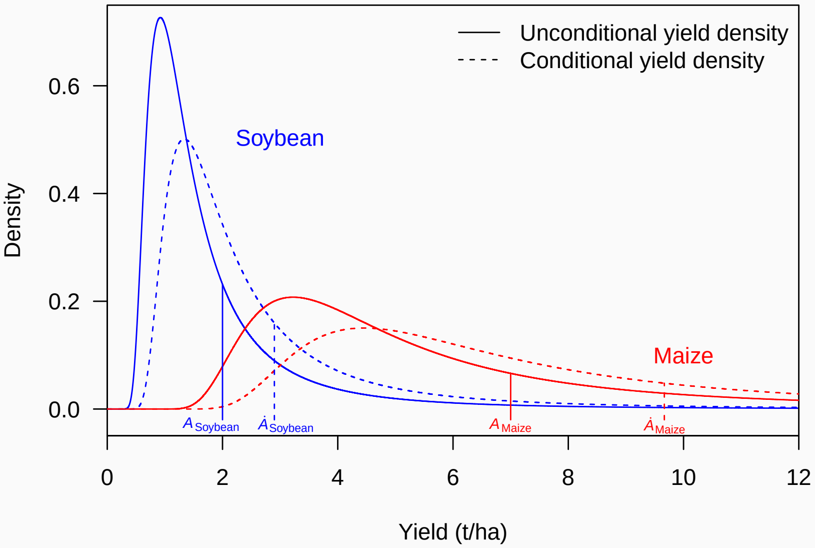
Fréchet-based modeling iii

- Production:

$$\begin{aligned} Q_k &= \bar{L}\pi_k \,\mathrm{E}\left[A_k(\omega) \,|\, R_k(\omega) \in \arg\max_{l \in \mathcal{K}} R_l(\omega) \right] \\ &= \bar{L} A_k \pi_k^{(\theta-1)/\theta}. \end{aligned}$$

The more a crop is planted, the lower its yield.

Illustration of land allocation with GEV



Comparison CET-Fréchet

CET

$$\pi_k = \frac{\gamma_k^{\eta} R_k^{1-\eta}}{\sum_{l \in \mathcal{K}} \gamma_l^{\eta} R_l^{1-\eta}},$$
$$Q_k = \bar{L} A_k \gamma_k^{\eta/(1-\eta)} \pi_k^{-\eta/(1-\eta)}.$$

Fréchet

$$\pi_k = \frac{R_k^{\theta}}{\sum_{l \in \mathcal{K}} R_l^{\theta}},$$
$$Q_k = \bar{L} A_k \pi_k^{(\theta-1)/\theta}.$$

Comparison CET-Fréchet

CET

$$\pi_k = \frac{\gamma_k^\eta R_k^{1-\eta}}{\sum_{l\in\mathcal{K}} \gamma_l^\eta R_l^{1-\eta}},$$
$$Q_k = \bar{L} A_k \gamma_k^{\eta/(1-\eta)} \pi_k^{-\eta/(1-\eta)}.$$

Fréchet

$$\pi_k = \frac{R_k^\theta}{\sum_{l\in\mathcal{K}} R_l^\theta},$$
$$Q_k = \bar{L} A_k \pi_k^{(\theta-1)/\theta}.$$

In relative deviation from benchmark ($\hat{x} \equiv x'/x$) with $\theta = 1-\eta$:

$$\hat{\pi}_k = \frac{\hat{R}_k^\theta}{\sum_{l\in\mathcal{K}} \pi_l \hat{R}_l^\theta},$$
$$\hat{Q}_k = \hat{A}_k \hat{\pi}_k^{(\theta-1)/\theta}.$$

Comparison CET-Fréchet

CET

$$\pi_k = \frac{\gamma_k^{\eta} R_k^{1-\eta}}{\sum_{l \in \mathcal{K}} \gamma_l^{\eta} R_l^{1-\eta}},$$
$$Q_k = \bar{L} A_k \gamma_k^{\eta/(1-\eta)} \pi_k^{-\eta/(1-\eta)}.$$

Fréchet

$$\pi_k = \frac{R_k^{\theta}}{\sum_{l \in \mathcal{K}} R_l^{\theta}},$$
$$Q_k = \bar{L} A_k \pi_k^{(\theta-1)/\theta}.$$

In relative deviation from benchmark ($\hat{x} \equiv x'/x$) with $\theta = 1 - \eta$:

$$\hat{\pi}_k = \frac{\hat{R}_k^{\theta}}{\sum_{l \in \mathcal{K}} \pi_l \hat{R}_l^{\theta}},$$
$$\hat{Q}_k = \hat{A}_k \hat{\pi}_k^{(\theta-1)/\theta}.$$

Same counterfactual changes for CET and Fréchet approaches except for land use changes.

Model for simulations

Key features of applied models

- (CES) Substitution between land and other inputs (possibly nested)
- Richer substitution patterns between land uses ($\Leftrightarrow$ nested CET functions)

Key features of applied models

- (CES) Substitution between land and other inputs (possibly nested)
- Richer substitution patterns between land uses ($\Leftrightarrow$ nested CET functions)

Can be done also with a Fréchet-based modeling.

Applied setup

- 1 non-land input, X_k, and $A_k^L(\omega) \sim$ multivariate Fréchet.

$$Q_k(\omega) = \left[\left(A_k^L(\omega) L(\omega)\right)^{(\sigma_k - 1)/\sigma_k} + \left(A_k^X X_k(\omega)\right)^{(\sigma_k - 1)/\sigma_k}\right]^{\sigma_k/(\sigma_k - 1)}$$

Applied setup

- 1 non-land input, X_k, and $A_k^L(\omega) \sim$ multivariate Fréchet.

$$Q_k(\omega) = \left[\left(A_k^L(\omega)\, L(\omega)\right)^{(\sigma_k - 1)/\sigma_k} + \left(A_k^X X_k(\omega)\right)^{(\sigma_k - 1)/\sigma_k}\right]^{\sigma_k/(\sigma_k - 1)}$$

- $R_k(\omega) = A_k^L(\omega)\, L(\omega) \left[p_k^{1-\sigma_k} - \left(A_k^X\right)^{\sigma_k - 1} w^{1-\sigma_k}\right]^{1/(1-\sigma_k)} \sim$ multivariate Fréchet

Applied setup

- 1 non-land input, X_k, and $A_k^L(\omega) \sim$ multivariate Fréchet.

$$Q_k(\omega) = \left[\left(A_k^L(\omega) L(\omega) \right)^{(\sigma_k - 1)/\sigma_k} + \left(A_k^X X_k(\omega) \right)^{(\sigma_k - 1)/\sigma_k} \right]^{\sigma_k/(\sigma_k - 1)}$$

- $R_k(\omega) = A_k^L(\omega) L(\omega) \left[p_k^{1-\sigma_k} - \left(A_k^X \right)^{\sigma_k - 1} w^{1-\sigma_k} \right]^{1/(1-\sigma_k)} \sim$ multivariate Fréchet
- Crops can be partitioned into N nests, $B_1, \ldots, B_N$. The land productivity vector has for CDF

$$F(a) = \exp\left\{ - \sum_{n=1}^{N} \left[\sum_{k \in B_n} \left(\frac{a_k}{\gamma A_k^L} \right)^{-\theta/(1-\rho_n)} \right]^{1-\rho_n} \right\} \quad \forall a \in \mathbb{R}_{>0}^K,$$

$0 \leq \rho_n < 1$ parameterizing the correlation within a nest ($\rho_n = 0 \Rightarrow$ statistical independence).

Applied model

Production

$$Q_k = \bar{L} A_k^L \pi_k^{(\theta-1)/\theta} \varpi_{k|n}^{\rho_n/\theta} \left(\frac{r_k}{p_k}\right)^{\sigma_k},$$

where

- $\pi_k = \Pi_n \varpi_{k|n}$: land use share of crop k
- $\Pi_n = \dfrac{\left(\sum_{l \in B_n} R_l^{\theta/(1-\rho_n)}\right)^{1-\rho_n}}{\sum_{m=1}^{N} \left(\sum_{l \in B_m} R_l^{\theta/(1-\rho_m)}\right)^{1-\rho_m}}$: land use share of nest n
- $\varpi_{k|n} = (R_k / R_{B_n})^{\theta/(1-\rho_n)}$: land use share of crop k in nest n

Own-price elasticity

$$\frac{\partial \ln Q_k}{\partial \ln p_k} = \underbrace{\frac{\theta}{\alpha_k^L}\left[(1-\pi_k) + \frac{\rho_n}{1-\rho_n}\left(1-\varpi_{k|n}\right)\right]}_{\text{Acreage}} \underbrace{\underbrace{-\frac{1-\pi_k}{\alpha_k^L}}_{\text{Composi.}} \underbrace{+\sigma_k\left(\frac{1}{\alpha_k^L}-1\right)}_{\text{Intensification}}}_{\text{Yield}},$$

Estimation

Model estimation

- To estimate: θ, ρ_n, and σ_k, and $\varpi_{k|n}$ if overlapping nests.
  - π_k calibrated to match observed land-use shares.

Model estimation

- To estimate: θ, ρ_n, and σ_k, and $\varpi_{k|n}$ if overlapping nests.
    - π_k calibrated to match observed land-use shares.
- A model based on GEV distributions represent a continuum of parcels which has been integrated
    - $\Rightarrow$ Not possible to estimate directly on farm-level information

Model estimation

- To estimate: θ, ρ_n, and σ_k, and $\varpi_{k|n}$ if overlapping nests.
  - π_k calibrated to match observed land-use shares.
- A model based on GEV distributions represent a continuum of parcels which has been integrated
  - $\Rightarrow$ Not possible to estimate directly on farm-level information
- Solution: 2-step estimation
  1. Estimation a multi-crop econometric model on farm-level data (Chambers and Just, 1989)
  2. Use a minimum distance estimation to calibrate the GEV-based model on the multi-crop econometric model behavior.

Multi-crop econometric model

- Translog profit function (Chambers and Just, 1989) with m crops and n inputs
- Input derived demands (Shephard's lemma) and output supply functions (Hotelling's lemma)
- Multi-crop profit function

$$\pi\left(\mathbf{p}, \mathbf{w}, z\right) = \max_{z_1,\ldots,z_m} \left\{ \sum_{i=1}^{m} \pi_i\left(p_i, \mathbf{w}, z_i\right) : \sum_{i=1}^{m} z_i = z \right\}$$

- Crop-specific profit function $(i = 1, \ldots, m)$

$$\pi_i = \exp\left\{ a_i + \alpha_i \ln p_i + \sum_{j=1}^{n} \beta_{ji} \ln w_j + \theta_i \ln z_i + \gamma_{0i} (\ln\ p_i)^2/2 + \sum_{j=1}^{n} \gamma_{ji} \ln p_i \ln w_j + \sum_{j,k=1}^{n} \phi_{kji} \ln w_k \ln w_j/2 \right\}$$

where $\phi_{kji} = \phi_{jki}$, and homogeneity constraints on parameters.

Data

- Farm-level data from European Commission Farm Accountancy Data Network (FADN) data set
- Unbalanced panel of European farmers, 1990–2015
  - 287,624 European farmers with 1,473,125 observations
- Data on m crop outputs and 3 inputs
- Inputs: land area, family labor, and other inputs

Minimum distance estimation

1. From the estimated Translog model calculate
   - Acreage and yield elasticities at the NUTS2 level: moments M
   - Variance-covariance matrix of M from the estimation: W.
2. Define from the GEV-based model the same moments: $M^s(\Theta)$
3. Find $\Theta = \{\theta, \rho_n, \sigma_k\}$ such that

$$\Theta = \arg\min_{\Theta} \left(M^s(\Theta) - M\right)' W^{-1} \left(M^s(\Theta) - M\right)$$

Local identification possible if # of independent moments $M \geq$ # of parameters to estimate Θ.

Conclusions

Conclusions

- This paper shows the relationships between land-use modeling based on CET and on GEV distributions.
- GEV distributions solve the pb of the non-respect of physical balances while implying minimal deviations from current modeling practices.
- Limit: GEV-based approach implies equality of conditional per hectare land rents across uses
  - But land-rent data are usually unavailable or available following strong assumptions (e.g., in GTAP),
  - Pbs where land use matters usually involve limited GE effects and could be addressed with PE models.

Thank you for your attention!

Christophe Gouel – christophe.gouel@inra.fr