Agricultural Land Use and Irrigation

Advanced Environmental Economics

In this assignment, you will learn how to estimate a dynamic structural model of agricultural land use in a water-scarce region. While some steps are guided, it is strongly recommended to read Scott (2013) for further details. This paper proposes and applies the Euler Conditional Choice Probability (ECCP) estimator in a similar setting. You can also check Kalouptsidi et al. (2021) for a more rigorous and general explanation of the method. The assignment is due on **June 30 2025**.

Setting

To analyze the economic impacts of climate change it is essential to understand the margins of adaptation in every sector. Agriculture is relatively more exposed to climate risk than other sectors due to its direct sensitivity to heat and precipitations. A large body of research focuses on estimating the effects of climate on agricultural yields (Schlenker et al., 2009). For instance, Burke and Emerick (2016) examine whether adaptation in the long run allows to mitigate the negative impact on yields in the short run. A mechanism of adaptation that is often studied is switching to crops or seeds that are more resistant to heat or dryness (Obolensky, 2024; Puy, 2024).

Many studies focus on non-irrigable areas to avoid modeling the irrigation choice. This is challenging due to the complex nature of water supply, varying irrigation policies over space and time, and limited data availability. However, irrigation is itself a margin of adaptation to climate change. Around a third of global agricultural production is supplied by irrigated water. The origin and source of irrigation water is also important since 60% of agricultural production from irrigated land is located in areas with high water stress. While irrigation can insure against climate risk in the short-run, it can lead to the depletion of water reserves if the extraction rates exceed the recharge rate. This motivates extending the existing literature on adaptation to climate change in water-scarce environments.

Datasets

The working sample comprises the southern Spain's region of Andalucía between 2017 and 2024. This region is relatively arid, yet it sustains an important agricultural production, to a large extent dependent on irrigation. There are 4 databases: a panel of land uses (`land_use_panel.csv`), spatial grid information (`sigpacGrid250.shp`), crop returns (`cropReturns.csv`), and weather variables (`weatherAnd.csv`).

Table 1: Panel of land uses

Var. name	**Description**
grid_id	Identifier of grid cell
prov	Province
year	Year
LU	Land use category
irrigation	Crop under irrigation (binary)

Table 2: Spatial grid information

Var. name	**Description**
grid_id	Identifier of grid cell
geometry	Geometry of the grid cell (polygon)

Table 3: Crop returns

Var. name	**Description**
prov	Province
year	Year
subcrop	Subcrop name (e.g. "apple")
crop	Crop name (e.g. "fruit")
price	Subcrop expected price (€/kg)
yield	Subcrop yield (kg/ha)
cost	Cost (€/ha)
subsidies	Subsidies (€/ha)
area	Total area of subcrop in the province (ha)

Table 4: Weather variables

Var. name	Description
prov	Province
year	Year
DD10	Degree days above 10ºC
DD30	Degree days above 30ºC
precipitation	Annual precipitation (mm)

Questions

1. Working with spatial data

   (a) Read the land use panel. Reclassify crops differentiating if they are irrigated or not.

   (b) Read the grid information dataset. Compute the area of the grid cells (in hectares).

   (c) Show in map the spatial distribution of cropland (irrigated, non-irrigated) in 2021 in the province of Almeria.

   (d) Descriptive statistics: compute total area by crop (average across all years), plot total area under irrigation over time (aggregate and by crop). Comment the results.

2. Model

   Individual landowners decide which crops to grow in their plot of land. Crops are defined by the land use category (arable land comprising annual crops, olive groves, nuts, and citrics) and whether they are irrigated or not[1]. The set of land use choices ($\mathcal{J}$) includes crops and the land use category "other", typically using land for pastures as an outside option.

   There is an individual field state (also called "controlled" state) with an observed part, k_{imt}, and a component unobserved by the econometrician, ε_{imt}. The individual state is deterministic, simply denoting the last period's use of land: $k_{imt} = a_{im,t-1}$. This captures the different switching costs in annual and perennial crops, but also costs of adopting irrigation technology. There is also an aggregate state $\omega_{mt} = (w_{mt}, \eta_{mt})$, where only variables in w_{mt} are observed by the econometrician.

   Assume that the flow profits for the landowner of grid cell i with crop j in province m

[1] For instance, irrigated annual crops are considered a different crop (j) than non-irrigated annual crops (j').

at year t are

$$\pi_j(s_{imt}) = \pi_j(k_{imt}, \omega_{mt}, \varepsilon_{ijmt}) = \theta_j + \theta_{kj} + \theta^R R^e_j(w_{mt}) + \xi_j(k_{imt}, \omega_{mt}) + \varepsilon_{ijmt},$$

where θ_j is an intercept for current land use j, θ_{kj} indicate switching costs from land use k to j, expected returns are

$$R^e_j(w_{mt}) = \begin{cases} p^e_{jmt} \cdot y^e_{jmt} - c_{jt} + subsidies_{jt}, & \text{if } j \text{ is a crop} \\ 350 & \text{if } j \text{ is "other"}, \end{cases}$$

where p^e_{jmt}, y^e_{jmt} are expected prices and expected yields, $c_{jt}, subsidies_{jt}$ are costs and total subsidies, both proportional to the land that is cultivated. Note that this implicitly assumes that there is perfect foresight of costs and subsidies for each plot, but prices and yields are predicted in every period.

Using Assumptions 1 and 2 from Scott (2013), we can define the value function and related concepts. The *value function* is given by the Bellman equation

$$V(s_{imt}) = \max_{j \in \mathcal{J}} \{\pi_j(s_{imt}) + \beta \mathbb{E}[V(s_{im,t+1}) | j, s_{imt}]\}.$$

The *ex-ante value function* is

$$V(k_{imt}, \omega_{mt}) := \int V(k_{imt}, \omega_{mt}, \varepsilon_{imt}) dF^{\varepsilon}(\varepsilon_{imt}),$$

and the *conditional value function* as

$$v_j(k_{imt}, \omega_{mt}) := \pi_j(k_{imt}, \omega_{mt}) + \beta \mathbb{E}[V(k_{im,t+1}, \omega_{m,t+1}) | j, k_{imt}, \omega_{mt}],$$

where $\pi_j(s_{imt}) = \pi_j(k_{imt}, \omega_{mt}) + \varepsilon_{imt}$. The agent's optimal policy is given by the *conditional choice probability (CCP)*:

$$p_j(k, \omega) = \int \mathbb{1}\{v_j(k, \omega) + \epsilon_j \geq v_l(k, \omega) + \varepsilon_j, \text{ for all } l \in \mathcal{J}\} dF^{\varepsilon}(\varepsilon).$$

From Arcidiacono and Miller (2011), for any (j, k, ω), there exists a real-valued function $\psi_j(\cdot)$ that satisfies the equality:

$$V(k, \omega) = v_j(k, \omega) + \psi_j(p(k, \omega)), \tag{1}$$

where $p(k, \omega)$ is the vector of stacked CCPs. This equation states that the ex-ante value function equals the value obtained by choosing any action j today and optimizing thereafter, plus a correction term.

3. Identification

We are going to derive the regression equation step by step, and obtain moment conditions. First, we express the correction term using the definition of the conditional value function. We simplify the notation of $\psi_j(p(k,\omega))$ to $\psi_j(k,\omega)$.

$$\begin{aligned}\psi_j(k,\omega) &= V(k,\omega) - \pi_j(k,\omega) - \beta\mathbb{E}[V(k',\omega')|j,k,\omega] \\ &= V(k,\omega) - \pi_j(k,\omega) - \beta V(k',\omega') - \beta e^v(j,k',\omega,\omega'),\end{aligned}$$

where $e^v(j,k,\omega,\omega^*) = \mathbb{E}_{\omega'|\omega}[V(k,\omega')] - V(k,\omega^*)$ is the *expectational error*. Notice that, since the field state in the next period is equal to the choice today ($k' = j$), we can omit the dependence of the last two terms from the field state in the next period. The notation simplifies to:

$$\psi_j(k,\omega) = V(k,\omega) - \pi_j(k,\omega) - \beta V(j,\omega') - \beta e^v(j,\omega,\omega'). \tag{2}$$

(a) Remove the current value function from equation (2) by taking the difference between two different crops j and a..

(b) Express the conditional value functions with respect to renewal action and take differences. Use the one-period finite dependence property to remove the continuation value from the difference in future value functions. *Hint: Crop switching is a renewal action: switching to crop J in the next period provides the same value in the future no matter which land use the agent is switching from.*

(c) Rearrange the resulting equation grouping the ψ terms in the LHS.

(d) Derive the regression equation using the logit inversion ($\psi_j(k,\omega) = \gamma - \ln(p_j(k,\omega))$). You can consider only $k = j$ and $J = a$.

(e) Discuss the identifying variation to recover the elasticity of land use to expected returns.

(f) Explain the endogeneity problem and propose instruments. To remove the fixed effects and obtain a moment condition, you can take first differences as in Scott (2013).

(g) Dynamic models are not generally identified without further restrictions. To recover the underlying parameters, θ_j, and θ_{kj}, from the estimated intercepts, impose that $\theta_{k,other}$ is independent from the field state k and normalize it to zero for some state k. Assume that switching costs are zero if costs do not switch, $\theta_{kk} = 0$ for all k.

4. Estimation

The ECCP is a two-step estimation method. First, CCP are estimated directly from the data and, secondly, they are introduced in the regression.

(a) Estimate the conditional choice probabilities using a frequency estimator for every year and province: $p_{jmt}(k) = \hat{p}_j(k, \omega)$. You will need to deal with zero probabilities.

(b) Read the datasets of crop returns and weather variables. Compute the expected yields as the fitted value of the regression:

$$\log(y_{cmt}) = \theta_{cm} + \theta_j W_{mt} + \theta_j t + \varepsilon^y_{cmt}, \tag{3}$$

for subcrop c belonging to crop j, W_{mt} includes degree days above 10ºC, degree days above 30ºC, precipitation, and interactions, and time trend t, like in Scott (2013). Compute and save the expected yields as the fitted values[2]. Aggregate the predicted yields of every subcrop up to the crop categories using the area by province and year.

(c) Obtain the dependent variable using $\beta = 0.9$.

(d) Estimate the elasticity by GMM or IV using the proposed instruments.

(e) Discuss and interpret the estimated elasticity. Since expected returns are built with aggregate provincial measures, the estimation can be quite noisy. Discuss how reducing the size of aggregate state can improve precision if finer data are available.

(f) Recover the estimated switching costs. For this, first you need to recover the intercepts from the regression equation once the elasticity is estimated after taking first differences.

(g) To examine if there is adaptation by irrigation, one can check whether realized yields are less sensitive to heat or precipitations from the yield model (3). However, even if irrigation is potentially beneficial, there could be substantial switching costs (e.g. fixed costs of installing an irrigation system). Examine the switching costs estimates and distinguish between crop switching costs and irrigation switching costs.

(h) Estimate the model i) with myopic agents and, ii) removing dynamics. Compare the elasticity estimates.

[2]While Scott (2013) takes the historical regional averages to form the yield predictions, here we will assume that farmers have perfect foresight about weather conditions. A more realistic approach should specify how farmers make predictions about future weather at the start of the growing season.

References

Arcidiacono, P. and R. A. Miller (2011). Conditional Choice Probability Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity. *Econometrica 79*(6), 1823–1867. Publisher: [Wiley, Econometric Society].

Arellano, M. and S. Bond (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies 58*(2), 277–297. Publisher: [Oxford University Press, Review of Economic Studies, Ltd.].

Burke, M. and K. Emerick (2016). Adaptation to Climate Change: Evidence from US Agriculture. *American Economic Journal: Economic Policy 8*(3), 106–140. Publisher: American Economic Association.

Kalouptsidi, M., P. T. Scott, and E. Souza-Rodrigues (2021, May). Linear IV regression estimators for structural dynamic discrete choice models. *Journal of Econometrics 222*(1, Part C), 778–804.

Magnac, T. and D. Thesmar (2002). Identifying Dynamic Discrete Decision Processes. *Econometrica 70*(2), 801–816. Publisher: [Wiley, Econometric Society].

Obolensky, M. (2024). Protect or Prepare? Crop Insurance and Adaptation in a Changing Climate. *Working Paper*.

Puy, T. D. (2024). An Empirical Model of Agricultural Subsidies with Environmental Externalities. *Working Paper*.

Schlenker, W., M. J. Roberts, and V. K. Smith (2009). Nonlinear Temperature Effects Indicate Severe Damages to U.S. Crop Yields under Climate Change. *Proceedings of the National Academy of Sciences of the United States of America 106*(37), 15594–15598. Publisher: National Academy of Sciences.

Scott, P. T. (2013, January). Dynamic Discrete Choice Estimation of Agricultural Land Use. *Working Paper*.