Linear Regression Equation from DDC

Coding Exercise
Environmental Economics

Zixuan

June 30, 2025

Section 1 Spatial Data

1. Reclassify crops based on land use and irrigation.

2. Calculate the size of grids.

3. Show the distribution of irrigated and non-irrigated croplands in Almeria in 2021 1.

4. Irrigated area by crop and aggregate over time 2.

5. Average area by crop 1. Average area by crop and irrigation 2.
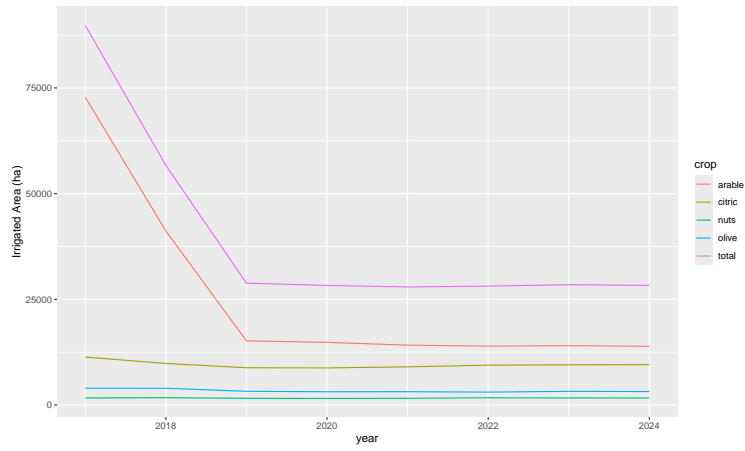
Section 2 Model

Instead of directly answering the questions, I develop the estimating equation in the following way (which answers the question indirectly).

crop	avg_area_ha
arable	149535.16
olive	9780.47
nuts	59607.03
citric	9535.94

Table 1: Average area by crop.
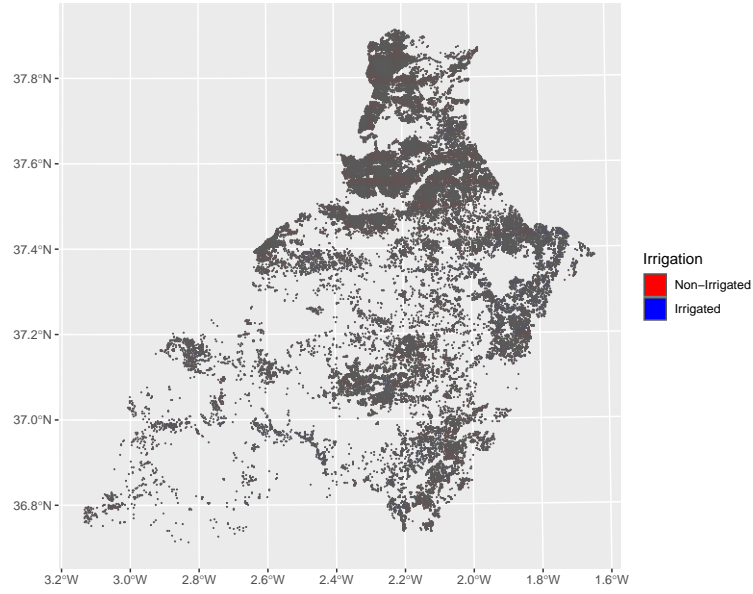
Figure 1: Distribution of irrigated and non-irrigated croplands in Almeria in 2021.

Figure 2: Irrigated area by crop and aggregate over time.

crop	irrigation	avg_area_ha
arable	0	124530.47
olive	1	3368.75
arable	1	25004.69
olive	0	6411.72
nuts	0	57960.16
nuts	1	1646.87
citric	1	9535.94

Table 2: Average area by crop and irrigation.

State   Usually, we write the observed state as s and unobserved state as ε. Here, the observed state has to two components: the individual state k and the aggregate state ω.

1. k evolves deterministically, i.e. $k_{t+1} = a_t$

2. ω evolves in the following way, $G(\omega_{t+1}|\omega_t, a_t) = G(\omega_{t+1}|\omega_t)$

Reward function and Value function   I denote the reward function as $u(k, \omega)$, and the continuation reward function as $\tilde{u}(k, \omega, a) = u(k, \omega, a) + \beta \mathbb{E}\left[\tilde{V}_{t+1}(k_{t+1}, \omega_{t+1})|k, \omega, a\right]$. Also value function is written as $V_t(k, \omega, \varepsilon)$ and the expected value function $\bar{V}_t(k, \omega)$.

Derivation   There is a nice relationship between the expected value function and the continuation reward function, which is

$$\bar{V}_t(k, \omega) = \tilde{u}(k, \omega, a) + \psi_a(p(k, \omega))$$

for any choice of action a. In the form that we are more familiar with by imposing the standard assumption on ε,

$$\bar{V}_t(k, \omega) = \tilde{u}(k, \omega, a) + \gamma - \Pr(a_t = a|k, \omega)$$

Now let us take another chioce $a_t = j$ which gives exactly the same stuff

$$\bar{V}_t(k, \omega) = \tilde{u}(k, \omega, a) + \psi_a(p(k, \omega))$$

Then we can take the difference between the two equations above,

$$\psi_j(p(k, \omega)) - \psi_a(p(k, \omega)) = \tilde{u}(k, \omega, a) - \tilde{u}(k, \omega, j)$$

Imposing the standard assumption on ε such that $\psi_a(p(k, \omega)) = \gamma - \log \Pr(a_t = a|k, \omega)$. The left hand side becomes

$$\log \frac{\Pr(a_t = j|k, \omega)}{\Pr(a_t = a|k, \omega)}$$

Look at the right hand side, how to express the $\tilde{u}(k, \omega, a)$ as an explicit function of observed variables and parameters? First, recall that individual state k evolves deterministically,

$$\tilde{u}(k, \omega, a) = u(k, \omega, a) + \beta \mathbb{E}\left[\bar{V}_{t+1}(a, \omega_{t+1})|k, \omega, a\right]$$

Then how to make use of renewal action? For example, we choose $a_t = a$ and $a_{t+1} = r$ OR $a_t = j$ and $a_{t+1} = r$. Since we the individual state is equal to the last period action, even if we are not switching crop, as long as we choose the same action, the individual state is the same (one period finite dependence?)

$$\mathbb{E}_{\omega_{t+1}|j, \omega} \bar{V}_{t+1}(r, \omega_{t+1}|j, \omega, r) = \mathbb{E}_{\omega_{t+1}|a, \omega} \bar{V}_{t+1}(r, \omega_{t+1}|a, \omega, r)$$

3

Recall in the case of Scott (2014), the renewal action is choosing crop, the state k is the number of years up to now that the land is not used for growing crops.

$$\mathbb{E}_{\omega_{t+1}}\bar{V}_{t+1}(k_{t+1}=2,\omega_{t+1}|k=1,\omega,a=\text{other}) \neq \mathbb{E}_{\omega_{t+1}}\bar{V}_{t+1}(k_{t+1}=3,\omega_{t+1}|k=2,\omega,a=\text{other})$$

But

$$\mathbb{E}_{\omega_{t+1}}\bar{V}_{t+1}(k_{t+1}=0,\omega_{t+1}|k=1,\omega,a=\text{crop}) = \mathbb{E}_{\omega_{t+1}}\bar{V}_{t+1}(k_{t+1}=0,\omega_{t+1}|k=2,\omega,a=\text{crop})$$

Back to the problem set, we have that by

$$\tilde{u}_t(k,\omega,a) = u_t(k,\omega,a) + \beta\mathbb{E}_{\omega_{t+1}}\bar{V}_{t+1}(a,\omega_{t+1}|k,\omega,a)$$

Similarly for $\tilde{u}_t(k,\omega,j)$, but of course

$$\mathbb{E}_{\omega_{t+1}|\omega}\bar{V}_{t+1}(a,\omega_{t+1}|k,\omega,a) \neq \mathbb{E}_{\omega_{t+1}|\omega}\bar{V}_{t+1}(j,\omega_{t+1}|k,\omega,j)$$

But if we look one period ahead, choosing the same action at $t+1$, then

$$\mathbb{E}_{\omega_{t+2}|\omega'}\bar{V}_{t+2}(r,\omega_{t+2}|a,\omega',r) = \bar{V}_{t+2|\omega'}(r,\omega_{t+2}|j,\omega',r)$$

Then the right hand side becomes

$$
\begin{aligned}
\tilde{u}_t(k,\omega,a) - \tilde{u}_t(k,\omega,j) &= u_t(k,\omega,a) - u_t(k,\omega,j) \\
&\quad + \beta\mathbb{E}_{\omega_{t+1}|\omega}\left[u_{t+1}(a,\omega_{t+1},r) - \ln\Pr_{t+1}(r|a,\omega_{t+1}) + \mathbb{E}_{\omega_{t+2}|\omega_{t+1}}\bar{V}_{t+2}(r,\omega_{t+2}|a,\omega_{t+1},r)\right] \\
&\quad - \beta\mathbb{E}_{\omega_{t+1}|\omega}\left[u_{t+1}(j,\omega_{t+1},r) - \ln\Pr_{t+1}(r|j,\omega_{t+1}) + \mathbb{E}_{\omega_{t+2}|\omega_{t+1}}\bar{V}_{t+2}(r,\omega_{t+2}|j,\omega_{t+1},r)\right] \\
&= u_t(k,\omega,a) - u_t(k,\omega,j) + \beta\mathbb{E}_{\omega_{t+1}|\omega}\left[u_{t+1}(a,\omega_{t+1},r) - u_{t+1}(j,\omega_{t+1},r)\right] \\
&\quad + \beta\mathbb{E}_{\omega_{t+1}|\omega}\left[\ln\Pr_{t+1}(r|a,\omega_{t+1}) - \ln\Pr_{t+1}(r|j,\omega_{t+1})\right]
\end{aligned}
$$

$$\tag{1}$$

Since we don't know the evolution fo aggregate state ω (we didn't even define it), by fixing one $\omega_{t+1} = \omega'$ and

$$
\begin{aligned}
\tilde{u}_t(k,\omega,a) - \tilde{u}_t(k,\omega,j) &= u_t(k,\omega,a) - u_t(k,\omega,j) \\
&\quad + \beta\left[u_{t+1}(a,\omega',r) - \ln\Pr_{t+1}(r|a,\omega') + \mathbb{E}_{\omega_{t+2}|\omega'}\bar{V}_{t+2}(r,\omega_{t+2}|a,\omega',r)\right] \\
&\quad - \beta\left[u_{t+1}(j,\omega',r) - \ln\Pr_{t+1}(r|j,\omega') + \mathbb{E}_{\omega_{t+2}|\omega_{t+1}}\bar{V}_{t+2}(r,\omega_{t+2}|j,\omega',r)\right] \\
&\quad + e_t(\omega',\omega,a) - e_t(\omega',\omega,j)
\end{aligned}
$$

$$\tag{2}$$

where $e_t(\omega', \omega, a) = \beta \mathbb{E}_{\omega_{t+1}|\omega} \bar{V}_{t+1}(a, \omega_{t+1}|k, \omega, a) - \beta \bar{V}_{t+1}(a, \omega'|k, \omega, a)$ which is the difference between the expecatation over $\omega_{t+1}|\omega$ and a random fixed $\omega_{t+1} = \omega'$.

Therefore the regression equation is

$$\log \frac{\Pr(a_t = j|k, \omega)}{\Pr(a_t = a|k, \omega)} + \beta \log \frac{\Pr_{t+1}(r|a, \omega')}{\Pr_{t+1}(r|j, \omega')} = u_t(k, \omega, a) - u_t(k, \omega, j)$$
$$+ \beta \left[u_{t+1}(a, \omega', r) - u_{t+1}(j, \omega', r) \right] + e_t(\omega', \omega, a) - e_t(\omega', \omega, j) \tag{3}$$

Plug in the functional form of the reward function,

$$u_t(k, \omega = (m, \eta), a) = \theta_a + \theta_{ka} + \theta_R R(m, a) + \xi(k, \omega, a)$$

Then the regression equation is

- $Y_{tk} = \log \frac{\Pr(a_t=j|k,\omega)}{\Pr(a_t=a|k,\omega)} + \beta \log \frac{\Pr_{t+1}(r|a,\omega')}{\Pr_{t+1}(r|j,\omega')}$

- $\theta_j - \theta_a + \theta_{kj} - \theta_{ka} + \beta(\theta_{jr} - \theta_{ar})$

- $\theta_R(R(m, a) - R(m, j))$

- $\xi(k, \omega, j) - \xi(k, \omega, a) + \beta(\xi_{t+1}(j, \omega', r) - \xi_{t+1}(a, \omega', r))$

- $e_t(\omega', \omega, a) - e_t(\omega', \omega, j)$

This is like the when we use market share data to estimate the parameters. In a market,

$$\log(s_j/s_0) = \alpha + \beta X_j + \xi_j$$

With multiple years of observations of the market.

$$\log(s_{jt}/s_{0t}) = \alpha + \beta X_{jt} + \xi_{jt}$$

if for a given j, ξ_{jt} is correlated across time, we should control for it $\xi_{jt} = \xi_j + \varepsilon_{jt}$. Then use the fixed effect estimator etc. Similarly here, the unit of observation is $k - j - a - r$. and fixed effect corresponds to.

Source of variation The identifying variation is cross sectional variation in $R(m, j) - R(m, a)$ for different $j - a$ pairs and across time variation $R_t(m, j) - R_t(m, a)$ over different t (becasue we need to take difference later!).

Endogeneity First, ignoring the fixed effect term, he terms xxx in the euqation are treated as the error term in the regression but one needs to justify that they are uncorrelated with $\Delta R(m, j, a)$ (across section endogeneity). For example, it may be the case the individual land $\xi(k, \omega, a)$ is uncorrelated with some variables m, and $R(m, j)$ is correlated with that m. Then we can use m as the instrument.

Secondly, to remove fixed effect, we need to take first difference, which introduces more endogeneity concerns (over time endogeneity). For example, the error term $\xi(k, \omega, a)$ is correlated with $R(m, j)$ over time. The first difference GMM estimator can be used to deal with the issue. Use lagged values of regressors x_{t-2} as instruments for the first difference of the regressors $\Delta x_t = x_t - x_{t-1}$.

Restriction and normalization

1. How many fixed effect can we estimate? $|k|^4$ fixed effects.

2. How many parameters do we have? $|k| + |k|^2$.

If we just fix one $j - a - r$, there are $|k|$ fixed effect, and $2|k| + 1$ parameters (θ_{kj},θ_{ka},the rest) to estimate. Some restrictions are needed.

1. we set θ_{ka} to be the same for all k. reduce by $|k| - 1$.

2. we assume non-switching cost is zero such that $\theta_{jj} = 0$. reduce by 1.

3. one more restriction is needed. therefore we set $\theta_{ka} = 0$ as well. reduce by 1.

But this is actually not necessary. Perhaps we have over identifification. But I will just leave it there as it is.

Section 3   Estimation

Observations filtering We restrict ourselves to the following two paths of actions, rather than any possible paths.

1. $k \to k \to a$

2. $k \to a \to a$

Therefore we have

- $Y_{tk} = \log \frac{\Pr(a_t = k | k, \omega)}{\Pr(a_t = a | k, \omega)} + \beta \log \frac{\Pr_{t+1}(a | k, \omega')}{\Pr_{t+1}(a | a, \omega')}$

- $\theta_k - \theta_a + \theta_{kk} - \theta_{ka} + \beta(\theta_{ka} - \theta_{aa})$

- $\theta_R(R(m, a) - R(m, k))$

- $\xi(k, \omega, k) - \xi(k, \omega, a) + \beta(\xi_{t+1}(k, \omega', a) - \xi_{t+1}(a, \omega', a))$

- $e_t(\omega', \omega, a) - e_t(\omega', \omega, k)$

Left-hand side variable To construct the left-hand side variable of the estimating equation, we need to estimate the choice probability first $\Pr(a_t = a|k, \omega)$. For each province-year-state-action pair, we use the frequency of action as the probability.

Right-hand side variable To construct the right-hand side variables, we estiamte the return by first predicting the yields and price, then plus subsidies and minus costs. The estimating equation is that we split the data into subcrop category, then we run for each category

`lm(log(yield) ~ DD30 * precipitation + DD10 * precipitation + year + prov )`

There seems to large differences between the predicted return and the actual return. Finer data maybe helpful as in Scott (2014) because in that paper, there are both state, country and land, three levels. While here we only have province and land. If there is a county level data on weather, we can better predict the return.

In the following estimation, I used the actual return.

Panel data estimation We first take difference to remove the fixed effect. Then we use the lagged value $x_{k,t-2}$ as the instrument for the first difference of the regressors $\Delta x_{kt} = x_{kt} - x_{k,t-1}$.

The esimating moment condition is

$$E[(\Delta y_{kt} - \beta \Delta x_{kt}) \cdot x_{k,t-2}] = 0$$

Recover fixed effect and switching cost parameters We use the following equations to recover the fixed effect α_i where $\bar{y}_i, \bar{x}_i$ is the time average. And i is the index for the unit (k-a pair)

$$\hat{\alpha}_i = \bar{y}_i - \beta \bar{x}_i$$

Recall that

$$\alpha_{ka} = \theta_k - \theta_a + \theta_{kk} - \theta_{ka} + \beta(\theta_{ka} - \theta_{aa}) = \theta_k - \theta_a - \theta_{ka} + \beta \theta_{ka}$$

1. step 1: pick any k_1 and $a_0 =$ other, we have $\alpha_{k_1 a_0} - \alpha_{a_0 a_0} = \theta_{k_1} - \theta_{a_0} + (\beta - 1)(\theta_{k_1 a_0} - \theta_{a_0 a_0}) = \theta_{k_1} - \theta_{a_0}$

2. step 2: pick any a_1 Then $\alpha_{k_1 a_1} - \alpha_{k_1 a_0} = \theta_{a_0} - \theta_{a_1} + (\beta - 1)(\theta_{k_1 a_1} - \theta_{k_1 a_0}) = \theta_{a_0} - \theta_{a_1} + (\beta - 1)\theta_{k_1 a_1}$ Since we can get any $\theta_{a_0} - \theta_{a_1}$ from step 1. Then we can automatically recover $\theta_{k_1 a_1}$

3. fixing k_1 and a_1 to be of the same crop but different irrigation will give irrigation switching cost. fixing them of differnt crops but same irrigation will give crop switching cost.

Myopic and static model

- For the myopic model, we set $\beta = 0$ and everthing else the same.

- For the static model which basically entirely removes state dependence. We need to recalculate the choice probability not conditional on state. T

$$\log(p_{jt}/p_{0t}) = \alpha + \beta X_{jt} + \xi_{jt} = \beta X_{jt} + \xi_j + \varepsilon_{jt}$$

Then use the GMM as above.

References

Scott, P. (2014). Dynamic discrete choice estimation of agricultural land use.