Linear Regression Equation from DDC

Coding Exercise
Environmental Economics

Zixuan

June 20, 2025

Question 1 Spatial Data

1.1 Land size

grid_id	prov	area
04_1	Almeria	6.25
04_2	Almeria	6.25
04_3	Almeria	6.25
04_4	Almeria	6.25
04_5	Almeria	6.25
04_6	Almeria	6.25
04_7	Almeria	6.25
04_8	Almeria	6.25
04_9	Almeria	6.25
04_10	Almeria	6.25

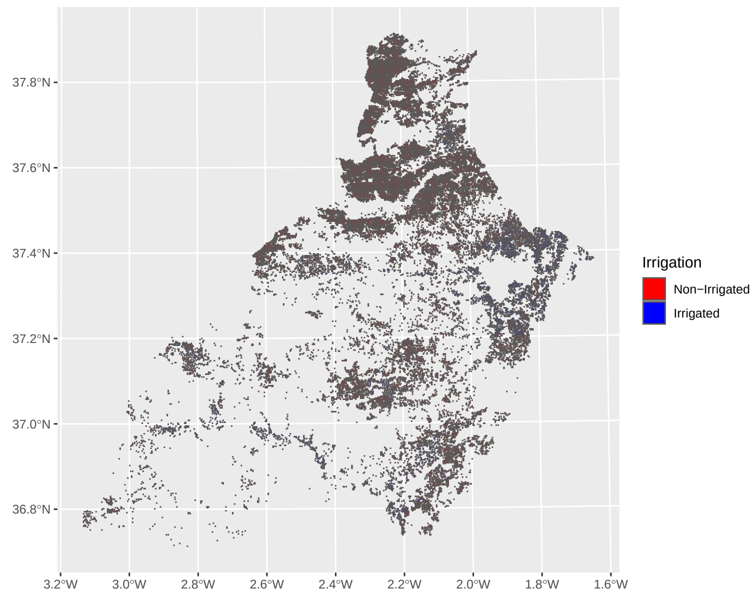
1.2 Cropland in Almeria

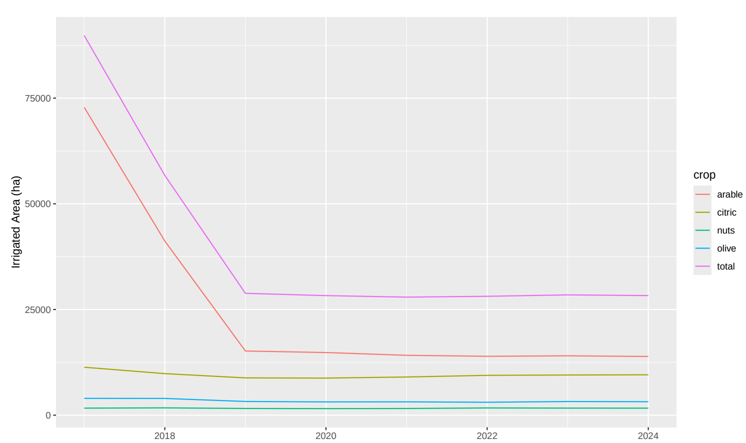
1.3 Cropland size under irrigation over time

Question 2 Model

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, $G(\omega_{t+1}|\omega_t, a_t) = G(\omega_{t+1}|\omega_t)$


37.8°N
37.6°N
37.4°N
37.2°N
37.0°N
36.8°N
3.2°W
3.0°W
2.8°W
2.6°W
2.4°W
2.2°W
2.0°W
1.8°W
1.6°W
Irrigation
Non−Irrigated
Irrigated


Irrigated Area (ha)
75000
50000
25000
0
2018
2020
2022
2024
year
crop
arable
citric
nuts
olive
total

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?)

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

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

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

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!!! Since here $|k| \geq$, it is guaranteed that $|k|^4 > |k| + |k|^2$.

Question 3 Estimation

References

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