AREC422 Notes

Youpei Yan

March 29, 2019

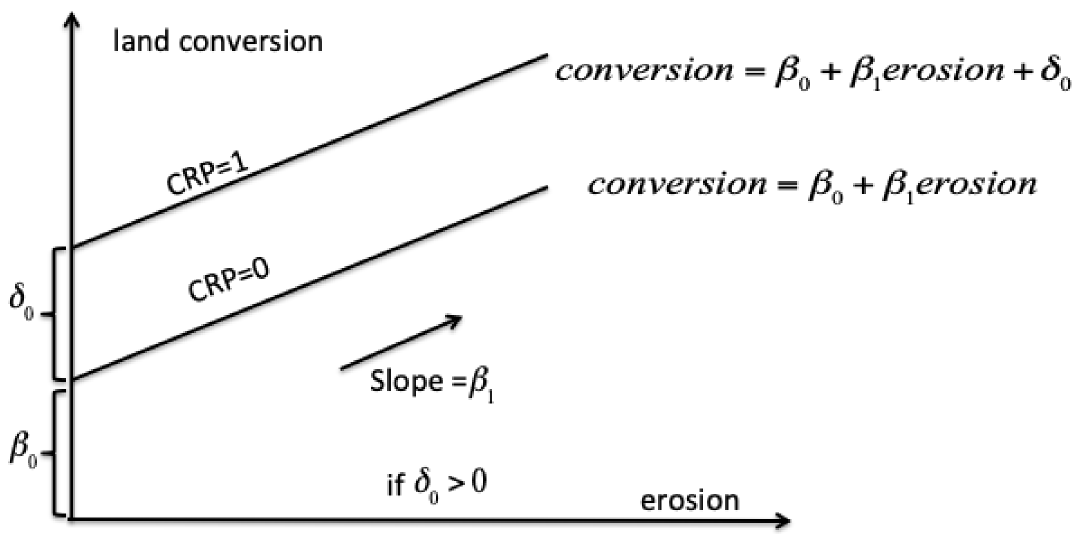
e.g.2: $land\_conversion = \beta_0 + \beta_1 erosion + \delta_0 CRP + u$. $CRP = 1$ if the conservation reserve program is in effect, $CRP = 0$ otherwise.

Mathematical interpretation:

$$\delta_0 = E(land\_conversion|CRP = 1, erosion) - E(land\_conversion|CRP = 0, erosion)$$

Word interpretation: Difference in the area of land conversion due to the conservation program, given the same erosion level.

Graphical interpretation:



Q1: why don't we include 2 dummy variables, say CRP and non_CRP to cover both cases?
A1: Avoid the perfect collinearity: $CRP + non\_CRP = 1$

Q2: how to test if the program's impact is significant?
A2: $H_0 : \delta_0 = 0$ vs. $H_1 : \delta_0 > 0$ (expect a positive impact). Then use t-value to perform the test.

This is also known as the **policy analysis** in the applied econometrics, as we have a **treatment group** ($CRP = 1$) and a **control group** ($CRP = 0$). The difference of the impact can be viewed as a policy evaluation.

e.g.3: $log(\widehat{pollution}) = \hat{\beta}_0 + \hat{\beta}_1 production + \hat{\beta}_2 abate + 0.297 industry\_zone$, where $industry\_zone$ is a dummy variable.

Recall the way to interpret the log-level model: we have the approximate impact = 29.7% if x increases by 1 unit, and the exact impact $= 100[e^{0.297} - 1]\% = 34.6\%$. Now, with the dummy x, the one can only increase by 1 unit, we interpret the coefficient in the following way:
"Holding *production* and *abatement* levels fixed", pollution will increase by around 29.7% (or 34.6% in the exact form) if the polluter is located inside the industrial zone.

* Ordinal Information Using Dummy Variables

If a categorical variable contains the information of more than 2 levels, how do we use this richer information in our estimation?

For instance, we have 4 levels of soil quality $soil\_qual = (A, B, C, D)$ or $(0, 1, 2, 3)$, with A or 0 as the best quality. How to incorporate this $soil\_qual$ into a model to explain the impact on $yield$? Regress directly if $soil\_qual = (0, 1, 2, 3)$? No.

A better approach is to create $3 = (4 - 1)$ dummy variables: $soil\_qi = 1$ if $soil\_qual = i$ and $soil\_qi = 0$ if $soil\_qual \neq i$ for $i = 0, 1, 2$

Note that we omitted the 4th level to avoid the collinearity issue. The omitted group is called the base group.

New model: $yield = \beta_0 + \delta_1 soil\_q0 + \delta_2 soil\_q1 + \delta_3 soil\_q2 + \beta_1 temp + \beta_2 ehdd + ... + u$

Interpret δ_2: Holding other factors fixed, $\hat{\delta}_2$ is the difference in $yield$ between a parcel of land with the level 1 quality and the level 3 quality (base group).

* Interactions with Dummy Variables

$y \sim x_1 + x_2 + D_1 \cdot D_2 + ...$
$y \sim x_1 + x_2 + D_1 \cdot x_1 + D_2 \cdot x_2 ...$

e.g.1.: $log(\widehat{afforestation}) = 0.321 - 0.11 flat\_land + 0.231 subsidy - 0.301 flat\_landcdotsubsidy$, where both $flat\_land$ (=1 if flat, =0 if sloped) and $subsidy$ (=1 if under a subsidy policy, =0 otherwise) are dummy variables. We can find the intercepts (when other factors are fixed) for the four different groups:

Intercept	$subsidy = 0$	$subsidy = 1$
$flat\_land = 0$	0.321	0.321+0.231
$flat\_land = 1$	0.321-0.110	0.321-0.110+0.231-0.301

We can compare groups' differences:
sloped land with subsidy provision vs. no subsidy provision: afforestation increases by 23.1%.
flat land with subsidy provision vs. no subsidy provision: afforestation decreases by 7%. etc.