

1. True or false: Under the unitary household model, increases in income from female members may have different effects on female consumption compared to an increase in income of the same amount increase from male members.

Nacy Qian proposes a test for whether households behave like an unitary model or a non-unitary model. She studies the impact of post-Mao reform in 1979 that increased income from cash crops including tea and orchard. Tea is considered predominantly female occupation whereas farming orchard is considered predominantly male occupation. In specific, she estimates regressions that are similar to

$$S_{iy} = \alpha + \beta TeaDum_i + \delta Post_y + \gamma (TeaDum_i \times Post_y) + \varepsilon_{iy}$$

where S_{iy} is the share of males in county i among those who were born in year y , $TeaDum_i$ is a dummy variable that takes value 1 if the region i grows tea and 0 otherwise. Similarly $Post_y$ is a dummy variable that takes value 1 if the cohort born in year y were born after the 1979 reforms. ε_{iy} is the error term.

2. What is the estimate γ called?
 - (a) Pre-post estimator
 - (b) Difference in Difference estimator
 - (c) Wald estimator
 - (d) IV (instrumental variable) estimator

Now load the data “tea_data.csv” in R (or any other statistical software of your liking. Before you begin, type these lines at the top of your R script:

```
install.packages("lfe")
```

```
library(lfe)
```

This will install a package that we will use later. The variable description of the data is as follows:

variable name	storage type	display format	value label	variable label
admin	float	%9.0g		County
biryr	int	%9.0g		Cohort: year of birth
han	float	%9.0g		Share of Han ethnicity
sex	float	%8.0g		Share of male
teasown	float	%9.0g		mu (unit of land) sown with tea
orch	float	%9.0g		size (in mu) of Orchard
cashcrop	float	%9.0g		size (in mu) of Cashcrop

3. First, lets understand the data structure. What describes the unit of observation in the data?
 - (a) An observation represents a county
 - (b) An observation represents a cohort of birth
 - (c) An observation represents a cohort of birth born in each county
 - (d) An observation represents a cohort of birth born in each county for each gender

Now, generate the variable **teaDum**, which takes value 1 if the any amount of tea is sown and 0 otherwise. Generate the variable **post**, which takes a value of 1 if the cohort is born on or after 1979. To be sure that you generated this variables correctly, the mean of **teaDum** is 0.209 and mean of **post** is 0.412. Now, also generate the interaction term.

4. Estimate γ from the data. Give its point estimate and the standard error. Is the point estimate significantly different from 0 at 95% level?
5. What is the right interpretation of the γ that you just estimated? (call your answer as x)

- (a) An increase in income from tea lowered the share of males by x
 - (b) The reform lowered the income from tea by x
 - (c) The reform lowered the share of males by x percentage points in counties that plant tea
 - (d) The reform lowered the share of males by x times 100 percent
6. Now generate variables `orchardDum` and `cashcropDum` in a way similar to `teaDum` (that is, they take value 1 if the crop is grown and 0 otherwise). Also, generate the interactions with `post`. Estimate the following regression

$$S_{iy} = \alpha + \beta_1 TeaDum_i + \beta_2 orchardDum_i + \beta_3 cashcropDum_i \\ + \gamma_1 (TeaDum_i \times Post_y) + \gamma_2 (orchardDum_i \times Post_y) + \gamma_3 (cashcropDum_i \times Post_y) \\ + \delta Post_y + \varepsilon_{iy}$$

In the following table, fill the point estimates and indicate whether the estimates are significant at 95% level.

7. Assuming that the policy did not change preferences for parents to have children of a specific gender, which of the following statements are necessary to reject the unitary model?
- (e) Keeping everything else the same, an increase in income from tea increased the share of girls
 - (f) Keeping everything else the same, share of male increased after the reform
 - (g) Keeping everything else the same, tea growing region had higher share of males for cohorts born before the reform
 - (h) Keeping everything else the same, an increase in income for orchard did not change the share of girls

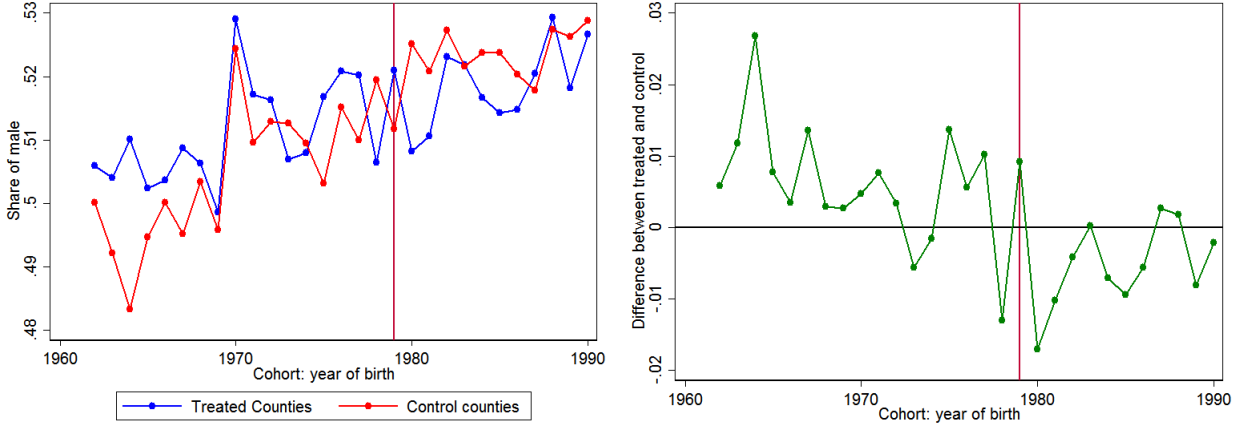
Now instead of grouping all the pre- and post- cohorts and grouping all the counties into two groups (those who grow and don't), let's exploit all the variation in the data. That is, let us estimate

$$S_{iy} = \gamma_1 (TeaDum_i \times Post_y) + \gamma_2 (orchardDum_i \times Post_y) + \gamma_3 (cashcropDum_i \times Post_y) \\ + \alpha + \psi_i + \rho_y + \varepsilon_{iy}$$

where ψ_i is a vector of dummies for each county and ρ_y is a vector of dummies for each year-of-birth cohort. (Hint: use the `felm` command in R instead of defining the vector of dummies)

8. What is the estimate of γ_1 from this model? Does this change the conclusion from previous regressions?
9. Now let's go back to the basic regression in question 2. What are the crucial assumptions to interpret γ as the causal impact of the policy?
- (a) The share of males in treated and control counties were the same before the reform
 - (b) The difference in share of males (share in treated – share in control) would have remained the same after the reform had the reform not taken place
 - (c) The share of males in treated and control counties would have been the same after the reform had the reform not taken place
 - (d) The reform reduced the share of males from the treated counties.

Examine the figure below



The figure on the left plots the average share of men in treated (blue) and control (red) counties for each birth cohort. The reform year is indicated by the vertical line. The figure on the right shows the difference between those two lines for each birth cohort.

10. From these figures, do you think the assumption(s) in question 9 is likely to hold?
- (a) Yes, the difference is positive before the reform and became negative afterwards
 - (b) No, the share of males were rising before the reform in both groups
 - (c) No, the difference in shares were falling before the reform
 - (d) Yes, the gap in share of males is narrowing down over time
11. Nancy uses the slope of the land (which determines whether the land is suitable for tea plantation) as an instrument for the area of land that planted tea. Let Z represent the slope of land and suppose that it takes values 1 if the land is suitable for tea after the reform and 0 otherwise (this is not the case, but we will assume this for simplicity). Table III shows the results. Which of the following comparison is column (4) showing
- (a) $E[S|Z = 1] - E[S|Z = 0]$, where S is the share of males
 - (b) $\frac{E[S|Z=1] - E[S|Z=0]}{E[TeaDum \times Post|Z=1] - E[TeaDum \times Post|Z=0]}$
 - (c) $(E[S|Z = 1] - E[S|Z = 0]) - (E[TeaDum \times Post|Z = 1] - E[TeaDum \times Post|Z = 0])$
 - (d) $E[TeaDum \times Post|Z = 1] - E[TeaDum \times Post|Z = 0]$
12. Lets continue to assume that the instrument is a dummy variable. What would be the instrumental variable estimate of γ as defined in 2?
- (a) $E[S|Z = 1] - E[S|Z = 0]$, where S is the share of males
 - (b) $\frac{E[S|Z=1] - E[S|Z=0]}{E[TeaDum \times Post|Z=1] - E[TeaDum \times Post|Z=0]}$
 - (c) $(E[S|Z = 1] - E[S|Z = 0]) - (E[TeaDum \times Post|Z = 1] - E[TeaDum \times Post|Z = 0])$
 - (d) $E[TeaDum \times Post|Z = 1] - E[TeaDum \times Post|Z = 0]$