Sex Ratio and the Price of Agricultural Crops in China

In this exercise, we consider the effect of a change in the price of agricultural goods whose production and cultivation are dominated by either men or women.

This exercise is based on: Qian, Nancy. 2008. "Missing Women and the Price of Tea in China: The Effect of Sex-Specific Earnings on Sex Imbalance." *Quarterly Journal of Economics* 123(3): 1251–85.

Our data come from China, where centrally planned production targets during the Maoist era led to changes in the prices of major staple crops. We focus here on tea, the production and cultivation of which required a large female labor force, as well as orchard fruits, for which the labor force was overwhelmingly male. We use price increases brought on by government policy change in 1979 as a proxy for increases in sex-specific income, and ask the following question: Do changes in sex-specific income alter the incentives for Chinese families to have children of one gender over another? The CSV data file, `chinawomen.csv`, contains the variables shown in the table below, with each observation representing a particular Chinese county in a given year. Note that `post` is an indicator variable that takes takes 1 in a year following the policy change and 0 in a year before the policy change.

Name	Description
`birpop`	Birth population in a given year
`biryr`	Year of cohort (birth year)
`cashcrop`	Amount of cash crops planted in county
`han`	Fraction ethnically Han
`orch`	Amount of orchard-type crops planted in county
`teasown`	Amount of tea sown in county
`sex`	Proportion of males in birth cohort
`post`	Indicator variable for introduction of price reforms

Question 1

We begin by examining sex ratios in the post-reform period (that is, the period after 1979) according to whether or not tea crops were sown in the region. Estimate the mean sex ratio in 1985, which we define as the proportion of male births, separately for tea-producing and non-tea-producing regions. Compute the 95% confidence interval for each estimate by assuming independence across counties within a year (We will maintain this assumption throughout this exercise). Furthermore, compute the difference-in-means between the two regions and its 95% confidence interval. Are sex ratios different across these regions? What assumption is required in order for us to interpret this difference as causal?

Question 2

Repeat the analysis in the previous question for subsequent years, i.e., 1980, 1981, 1982, ..., 1990. Create a graph which plots the difference-in-means estimates and their 95% confidence intervals against years. Give a substantive interpretation of the plot.

Question 3

Next, we compare tea-producing and orchard-producing regions before the policy enactment. Specifically, we examine the sex ratio and the proportion of Han Chinese in 1978. Estimate the mean difference, its standard

error, and 95% confidence intervals for each of these measures between the two regions. What do the results imply about the interpretation of the results given in Question~1?

Question 4

Repeat the analysis for the sex ratio in the previous question for each year before the reform, i.e., from 1962 until 1978. Create a graph which plots the difference-in-means estimates between the two regions and their 95% confidence intervals against years. Give a substantive interpretation of the plot.

Question 5

We will adopt the difference-in-differences design by comparing the sex ratio in 1978 (right before the reform) with that in 1980 (right after the reform). Focus on a subset of counties that do not have missing observations in these two years. Compute the difference-in-differences estimate and its 95% confidence interval. Note that we assume independence across counties but account for possible dependence across years within each county. Then, the variance of the difference-in-differences estimate is given by:

$$\mathbb{V}\{(\overline{Y}_{tea,after} - \overline{Y}_{tea,before}) - (\overline{Y}_{orchard,after} - \overline{Y}_{orchard,before})\}$$
$$= \mathbb{V}(\overline{Y}_{tea,after} - \overline{Y}_{tea,before}) + \mathbb{V}(\overline{Y}_{orchard,after} - \overline{Y}_{orchard,before})$$

where dependence across years is given by:

$$\mathbb{V}(\overline{Y}_{tea,after} - \overline{Y}_{tea,before}) = \mathbb{V}(\overline{Y}_{tea,after}) - 2\mathrm{Cov}(\overline{Y}_{tea,after}, \overline{Y}_{tea,before}) + \mathbb{V}(\overline{Y}_{tea,before})$$

A similar formula can be given for orchard-producing regions. What substantive assumptions does the difference-in-differences design require? Give a substantive interpretation of the results.