

اقتصادسنجی دوره فرعی، ۴۴۷۲۰

ترم دوم سال تحصیلی ۱۳۹۸-۹۹

## تمرین سری دهم: تفاضل در تفاضل

موعده تحویل: سه شنبه ۲۷ خرداد ۱۳۹۹، ساعت ۱۲:۳۰

نکته: لازم است پاسخ تمرین را به صورت یک فایل PDF در سامانه درس افزار آپلود نمایید.

---

1. You hear about a program that started providing technical assistance and general business training to farmers in ten of Ghana's cocoa growing regions in 1998 and want to know if it was successful at increasing yields.

(a) You have data on cocoa yields for those districts in 1995 and 2000. In 1995 and 2000, the average yield was 290 kg/ha and 400 kg/ha respectively. Is this difference of 110 kg/ha a good estimate of the program's effect? Why or why not? If you think it is not a good estimate, give specific examples of what could be distorting this estimate? Is there a reason to believe it is either an over-or under-estimate of the program's effect?

(b) You find that in the eleven districts where the program did not operate, average yields in 2000 were 320 kg/ha. Is the difference in average yields between districts with and without the program in (80 kg/ha) a good estimate of the program's effect? Why or why not? Which, if any, of the problems that you identified in part (a) does it solve? Does this estimate suffer from any new problems?

(c) Recognizing that you might be able to do better, you find data from 1995 for the districts without the program. Average yields in these districts were 280 kg/ha. What is the difference-in-differences estimate of the program's effect? How does it solve some of the problems you identified in parts (a) and (b)?

(d) Why might you still be concerned about the validity of this estimate of the program's effect? Suppose you were able to get yield data from the same districts for 1985 and 1990. How could you use this data to critically assess the estimate you calculated in (c)? What other data might help you improve the quality of your estimate?

(e) Instead of using levels of yield, run the DD estimates with natural logarithm of yield. What do you conclude? What is the identification assumption when you use logs? Is it different from the one you discussed in (d)?

2. The ministry of health has built about 100 health centers in some villages. You have data on health outcomes in a sample of these villages and also other villages that did not get the health centers. For individual  $i$  living in village  $v$  define  $Y_{ivt}(0)$  to be the health outcome in year  $t$  if no health center was built in the village and  $Y_{ivt}(1)$  to be the health outcome in year  $t$  if a health center is built in the village. Individuals are observed for two periods  $t \in \{0,1\}$  and the treatment was carried out in  $t = 1$ .

(a) Is it possible to observe both  $Y_{ivt}(0)$  and  $Y_{ivt}(1)$ ? Explain.

Now assume that

$$Y_{ivt}(0) = \alpha + \beta d_t + \gamma T_v + \epsilon_{ivt}$$

$$Y_{ivt}(1) = Y_{ivt}(0) + \tau$$

(b) Derive the DD estimator with a double difference and in a regression.

(c) In the above part state the assumptions that will ensure the estimator will be a consistent estimator for  $\tau$ .

(d) Discuss what happens if we define potential outcomes as

$$Y_{ivt}(0) = \alpha_i + \beta d_t + \gamma T_v + \epsilon_{ivt}$$

$$Y_{ivt}(1) = Y_{ivt}(0) + \tau$$