ECO3121 Problem Set 4

Xue Zhongkai (122090636)

Question 1

1. As $Z_{it} = Z_t$, the variable Z_i is the same for the same land parcel but specific for the time, hence it could be a variable like **climate characteristic for the year or specific policy at the time**.

   Estimation method:

   - Method I: Introduce a fixed-effect variable α_t, re-write the model as

     $Crop_{it} = \beta_0 + \beta_1 X_{it} + \alpha_t + \gamma_t + u_{it}$

     Then we substract the entity average of time from it to remove the land-fixed effect:

     $Crop_{it} - \bar{Crop}_{it} = \beta_1(X_{it} - \bar{X}_{it}) + \alpha_t - \bar{\alpha}_t + \gamma_t - \bar{\gamma}_t + u_{it} - \bar{u_{it}}$

     which is the demeaned equation:

     $\tilde{Crop}_{it} = \beta_1 \tilde{X}_{it} + \tilde{u_{it}}$

     Now we could estimate coefficients by regressing $\tilde{Crop}_{it}$ on $\tilde{X}_{it}$.

   - Method II: Set the first year ($t = 1$) as the base, and re-write the model in binary regression form:

     $Crop_{it} = \beta_0 + \beta_1 X_{it} + \sum_{i=2}^{N} \alpha_t D_t + \gamma_t + u_{it}$

     Here $\sum_{t=2}^{N} \alpha_t D_t$ is the sum for time-fixed effect, and $D_t = 1$ for the tth year but 0 for other times.

     Finally we could estimate the coefficients by OLS.

2. As $Z_{it} = Z_i$, the variable Z_t is the same for all years but specific for land parcels, hence it could be a variable like **quality, position or size of the land parcels**.

   Estimation method:

   - Method I: Introduce a fixed-effect variable α_i, re-write the model as

     $Crop_{it} = \beta_0 + \beta_1 X_{it} + \alpha_i + \gamma_t + u_{it}$

     Then we substract the entity average of time from it to remove the land-fixed effect:

     $Crop_{it} - \bar{Crop}_{it} = \beta_1(X_{it} - \bar{X}_{it}) + \gamma_t - \bar{\gamma}_t + u_{it} - \bar{u_{it}}$

     which is the demeaned equation:

     $\tilde{Crop}_{it} = \beta_1 \tilde{X}_{it} + \tilde{\gamma}_t + \tilde{u_{it}}$

     Now we could estimate coefficients by regressing $\tilde{Crop}_{it}$ on $\tilde{X}_{it}$ and $\tilde{\gamma}_t$.

   - Method II: Set the first land parcel ($i = 1$) as the base, and re-write the model in binary regression form:

     $Crop_{it} = \beta_0 + \beta_1 X_{it} + \sum_{i=2}^{N} \alpha_i D_i + \gamma_t + u_{it}$

     Here $\sum_{i=2}^{N} \alpha_i D_i$ is the sum for land-fixed effect, and $D_i = 1$ for the ith land parcel but 0 for others.

     Finally we could estimate the coefficients by OLS.

Question 2

1. We could write difference-in-difference estimator as

   $\text{DiD} = (Y_i^1|_{D_i=1} - Y_i^1|_{D_i=0}) - (Y_i^0|_{D_i=1} - Y_i^0|_{D_i=0})$

   $Y_i^1|_{D_i=1}$: The crop of the land parcel in Gansu province, after the attack.

   $Y_i^0|_{D_i=1}$: The crop of the land parcel in Gansu province, before the attack.

   $Y_i^1|_{D_i=0}$: The crop of the land parcel in non-Gansu province, after the attack.

   $Y_i^0|_{D_i=0}$: The crop of the land parcel in non-Gansu province, before the attack.
2. The assumption of parallel trends: In the absence of the attack (free fertilizer), the average change in the crop would have been the same for the land parcel within and without Gansu province.
3. Under this circumstance, the DiD estimator will over-estimate the effect of the treatment (free fertilizer).

   Farmers in Gansu province working harder but not taken into account of the regression, will lead to a typical Omitted Variable Bias (OVB). In other words, the increase in crops results from not only the treatment observed (free fertilizer), but also the omitted variable (farmers working harder). Thus the effect of the treatment will be over-estimated.

Question 3

Here are a summary of the scenario:

Location	Treatment
New Jersey	Increase: $4.25 to $5.05
Pennsylvania	(Control Group)

1. Comparing New Jersey vs. PA after the change in the minimum wage may not be a good estimation, because there could be many other factors influencing the employment, as the two states vary in aspects like economic conditions, taxation policy and consumer preferences. Only if these differences are controlled will we assure whether the increase in employment results from the minimun wages.
2. Here is the regression with entity and time fixed effect:

   $Empl_{it} = \beta_0 + \beta_1 Treatment_t + \beta_2 Group_i + \beta_3 (Treatment_t \times Group_i) + \gamma X_{it} + \epsilon_{it}$

   where X_{it} is the set of control variables.
3. The assumption of parallel trends: In the absence of the treatment (increase in minimum wage), the average change in the employment would have been the same in New Jersey and PA.
4. First we import the `.csv` file:

   ```
   import delimited "/Users/kevinshuey/Github/coursework/eco3121/as4/DID_Example.csv"
   ```

   Then we do some pre-process on some `str` data:

```
gen temp_empl = real(Empl)
gen temp_soda = real(Soda)
gen temp_fries = real(Fries)
```

Finally we perform the regression:

```
. reg temp_empl Group Treatment Group#Treatment COwned HoursOpening /*
      */  temp_soda temp_fries PA1 Shore NorthJ CentralJ
```

I believe the ordinary `chain` is meaningless, thus I didn't include it. Also I omitted `PA2`, `SouthJ` in case of collinearity.

And here is the result:

Source	SS	df	MS	Number of obs	=	757
				F(11, 745)	=	8.47
Model	5553.1004	11	504.827309	Prob > F	=	0.0000
Residual	44416.1519	745	59.6189959	R-squared	=	0.1111
				Adj R-squared	=	0.0980
Total	49969.2523	756	66.0968946	Root MSE	=	7.7213

temp_empl	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Group	-5.306838	1.280375	-4.14	0.000	-7.82041	-2.793265
Treatment	2.694839	.679795	3.96	0.000	1.360297	4.029381
Group#Treatment						
0 1	-4.140921	1.421972	-2.91	0.004	-6.93247	-1.349371
1 0	0	(omitted)				
1 1	0	(omitted)				
COwned	-.9494139	.6110438	-1.55	0.121	-2.148987	.2501588
HoursOpening	.5096475	.0705747	7.22	0.000	.3710986	.6481965
temp_soda	.6660348	4.217945	0.16	0.875	-7.614439	8.946509
temp_fries	-3.829184	3.667151	-1.04	0.297	-11.02836	3.369995
PA1	-1.811641	1.30443	-1.39	0.165	-4.372437	.7491551
Shore	1.158525	1.046825	1.11	0.269	-.8965524	3.213602
NorthJ	2.979691	.813729	3.66	0.000	1.382216	4.577165
CentralJ	1.202135	.9695933	1.24	0.215	-.7013259	3.105595
_cons	6.857323	3.748683	1.83	0.068	-.5019166	14.21656

5. With the assumption of parallel trends, the employment trends in New Jersey and Pennsylvania were similar prior to the policy intervention. After the implementation of the policy, any changes in the difference between the two states can be attributed to the increase in the minimum wage. By comparing these differences i.e. DiD, we can control for other factors hence get an unbiased estimation.