Applied (Financial) Econometrics

Empirical exercise DiD

Monique de Haan
(monique.dehaan@uva.nl)

Introduction

- In this emprical exercise we are going to replicate part of the results in Card & Krueger (1994)
- C&K investigate effect of an increase in the minimum wage from $4.25 to $ 5.05 in New Jersey on April 1, 1992.
- The dataset *did.dta* contains the data used in Card & Krueger (1994)
- Data set contains information on 410 fast food restaurants in New Jersey (NJ) and Pennsylvania (PA) before and after an increase in minimum wages in New Jersey on 1 April 1992.
- A description of all variables can be found in the file *did.pdf*.

Introduction

TABLE 3—AVERAGE EMPLOYMENT PER STORE BEFORE AND AFTER THE RISE IN NEW JERSEY MINIMUM WAGE

	Stores by state			Stores in New Jersey[a]			Differences within NJ[b]	
Variable	PA (i)	NJ (ii)	Difference, NJ – PA (iii)	Wage = \$4.25 (iv)	Wage = \$4.26–\$4.99 (v)	Wage ≥ \$5.00 (vi)	Low–high (vii)	Midrange–high (viii)
1. FTE employment before, all available observations	23.33 (1.35)	20.44 (0.51)	–2.89 (1.44)	19.56 (0.77)	20.08 (0.84)	22.25 (1.14)	–2.69 (1.37)	–2.17 (1.41)
2. FTE employment after, all available observations	21.17 (0.94)	21.03 (0.52)	–0.14 (1.07)	20.88 (1.01)	20.96 (0.76)	20.21 (1.03)	0.67 (1.44)	0.75 (1.27)
3. Change in mean FTE employment	–2.16 (1.25)	0.59 (0.54)	2.76 (1.36)	1.32 (0.95)	0.87 (0.84)	–2.04 (1.14)	3.36 (1.48)	2.91 (1.41)
4. Change in mean FTE employment, balanced sample of stores[c]	–2.28 (1.25)	0.47 (0.48)	2.75 (1.34)	1.21 (0.82)	0.71 (0.69)	–2.16 (1.01)	3.36 (1.30)	2.87 (1.22)
5. Change in mean FTE employment, setting FTE at temporarily closed stores to 0[d]	–2.28 (1.25)	0.23 (0.49)	2.51 (1.35)	0.90 (0.87)	0.49 (0.69)	–2.39 (1.02)	3.29 (1.34)	2.88 (1.23)

Part 1

```
. generate emptot1=emppt*0.5+empft+nmgrs
(12 missing values generated)

. generate emptot2=emppt2*0.5+empft2+nmgrs2
(14 missing values generated)

. regress emptot1 state, robust

Linear regression                               Number of obs     =        398
                                                F(1, 396)         =       4.04
                                                Prob > F          =     0.0451
                                                R-squared         =     0.0138
                                                Root MSE          =     9.6947

------------------------------------------------------------------------------
             |               Robust
     emptot1 | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       state |  -2.891761   1.438687    -2.01   0.045    -5.720179   -.0633421
       _cons |   23.33117   1.345732    17.34   0.000      20.6855    25.97684
------------------------------------------------------------------------------
```

Part 1

```
. generate emptot1=emppt*0.5+empft+nmgrs
(12 missing values generated)

. generate emptot2=emppt2*0.5+empft2+nmgrs2
(14 missing values generated)

. regress emptot1 state, robust

Linear regression                               Number of obs     =        398
                                                F(1, 396)         =       4.04
                                                Prob > F          =     0.0451
                                                R-squared         =     0.0138
                                                Root MSE          =     9.6947
```

emptot1	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
state	-2.891761	1.438687	-2.01	0.045	-5.720179	-.0633421
_cons	23.33117	1.345732	17.34	0.000	20.6855	25.97684

- the constant term gives the average pre-treatment employment in PA
- the coefficient on state gives the pre-treatment difference between NJ-PA

Part 2

```
. generate statealt=1-state

. regress emptot1 statealt, robust

Linear regression                               Number of obs     =        398
                                                F(1, 396)         =       4.04
                                                Prob > F          =     0.0451
                                                R-squared         =     0.0138
                                                Root MSE          =     9.6947

------------------------------------------------------------------------------
             |               Robust
     emptot1 | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
    statealt |   2.891761   1.438687     2.01   0.045     .0633421    5.720179
       _cons |   20.43941   .5087483    40.18   0.000     19.43922    21.43959
------------------------------------------------------------------------------
```

Part 2

```
. generate statealt=1-state

. regress emptot1 statealt, robust

Linear regression                               Number of obs     =        398
                                                F(1, 396)         =       4.04
                                                Prob > F          =     0.0451
                                                R-squared         =     0.0138
                                                Root MSE          =     9.6947
```

emptot1	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
statealt	2.891761	1.438687	2.01	0.045	.0633421	5.720179
_cons	20.43941	.5087483	40.18	0.000	19.43922	21.43959

- the constant term gives the average pre-treatment employment in NJ

Part 3

```
. regress emptot2 state, robust

Linear regression                               Number of obs     =        396
                                                F(1, 394)         =       0.02
                                                Prob > F          =     0.8977
                                                R-squared         =     0.0000
                                                Root MSE          =     9.1058
```

emptot2	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
state	-.1381549	1.074157	-0.13	0.898	-2.24995	1.973641
_cons	21.16558	.9394517	22.53	0.000	19.31862	23.01255

Part 3

```
. regress emptot2 state, robust

Linear regression                               Number of obs     =        396
                                                F(1, 394)         =       0.02
                                                Prob > F          =     0.8977
                                                R-squared         =     0.0000
                                                Root MSE          =     9.1058

------------------------------------------------------------------------------
             |               Robust
     emptot2 | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       state |  -.1381549   1.074157    -0.13   0.898     -2.24995    1.973641
       _cons |   21.16558   .9394517    22.53   0.000     19.31862    23.01255
------------------------------------------------------------------------------
```

- the constant term gives the average post-treatment employment in PA
- the coefficient on state gives the post-treatment difference between NJ-PA

Part 3

```
. regress emptot2 statealt, robust

Linear regression                               Number of obs     =        396
                                                F(1, 394)         =       0.02
                                                Prob > F          =     0.8977
                                                R-squared         =     0.0000
                                                Root MSE          =     9.1058
```

emptot2	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
statealt	.1381549	1.074157	0.13	0.898	-1.973641	2.24995
_cons	21.02743	.5208101	40.37	0.000	20.00352	22.05134

Part 3

```
. regress emptot2 stateal t, robust
```

Linear regression				Number of obs	=	396
				F(1, 394)	=	0.02
				Prob > F	=	0.8977
				R-squared	=	0.0000
				Root MSE	=	9.1058

emptot2	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
statealt	.1381549	1.074157	0.13	0.898	-1.973641	2.24995
_cons	21.02743	.5208101	40.37	0.000	20.00352	22.05134

- the constant term gives the average post-treatment employment in NJ

Part 4

```
. generate emptotd=emptot2-emptot1
(26 missing values generated)

. regress emptotd state, robust

Linear regression                               Number of obs     =        384
                                                F(1, 382)         =       4.23
                                                Prob > F          =     0.0405
                                                R-squared         =     0.0146
                                                Root MSE          =     8.9678
```

emptotd	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
state	2.75	1.337725	2.06	0.040	.1197732	5.380227
_cons	-2.283333	1.24814	-1.83	0.068	-4.737419	.1707523

Part 4

```
. generate emptotd=emptot2-emptot1
(26 missing values generated)

. regress emptotd state, robust

Linear regression                               Number of obs     =        384
                                                F(1, 382)         =       4.23
                                                Prob > F          =     0.0405
                                                R-squared         =     0.0146
                                                Root MSE          =     8.9678

------------------------------------------------------------------------------
             |               Robust
     emptotd | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       state |       2.75   1.337725     2.06   0.040     .1197732    5.380227
       _cons |  -2.283333    1.24814    -1.83   0.068    -4.737419    .1707523
------------------------------------------------------------------------------
```

- the constant term gives the trend in employment in PA
- the coefficient on state gives the difference-in-differences estimate (row 3, col 3 CK94 Table 3)

Part 5

```
. keep if emptot1!=. & emptot2!=.
(26 observations deleted)

. keep restaurant_id state emptot1 emptot2

. reshape long emptot, i( restaurant_id) j(time)
(j = 1 2)

Data                                  Wide   ->   Long
-----------------------------------------------------------------------------
Number of observations                 384   ->   768
Number of variables                      4   ->   4
j variable (2 values)                        ->   time
xij variables:
                           emptot1 emptot2   ->   emptot
-----------------------------------------------------------------------------
```

Part 5

	state	restaurant~d	time	emptot
1	0	1	1	40.5
2	0	1	2	24
3	0	2	1	13.75
4	0	2	2	11.5
5	0	3	1	8.5
6	0	3	2	10.5
7	0	4	1	34
8	0	4	2	20
9	0	5	1	24
10	0	5	2	35.5

Part 6

```
. xtset restaurant_id

Panel variable: restaurant_id (balanced)

. gen treated=0

. replace treated=1 if state==1 & time==2
(309 real changes made)

. xi: xtreg emptot treated i.time, fe robust
i.time            _Itime_1-2          (naturally coded; _Itime_1 omitted)

Fixed-effects (within) regression               Number of obs     =        768
Group variable: restaurant~d                    Number of groups  =        384

R-squared:                                      Obs per group:
     Within  = 0.0147                                         min =          2
     Between = 0.0055                                         avg =        2.0
     Overall = 0.0000                                         max =          2

                                                F(2,383)          =       2.14
corr(u_i, Xb) = -0.0978                         Prob > F          =     0.1187

                        (Std. err. adjusted for 384 clusters in restaurant_id)
------------------------------------------------------------------------------
             |               Robust
      emptot | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     treated |       2.75   1.337723     2.06   0.040     .1197995    5.380201
    _Itime_2 |  -2.283333   1.248138    -1.83   0.068    -4.737394    .1707278
       _cons |   21.00664   .2288166    91.81   0.000     20.55675    21.45653
-------------+----------------------------------------------------------------
     sigma_u |  8.4585732
     sigma_e |  6.3411612
         rho |  .64020113   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Part 7

```
. xi: regress emptot treated i.state i.time, robust
i.state           _Istate_0-1        (naturally coded; _Istate_0 omitted)
i.time            _Itime_1-2         (naturally coded; _Itime_1 omitted)

Linear regression                               Number of obs     =        768
                                                F(3, 764)         =       1.35
                                                Prob > F          =     0.2557
                                                R-squared         =     0.0076
                                                Root MSE          =     9.5113
```

emptot	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
treated	2.75	1.842828	1.49	0.136	-.8676072	6.367607
_Istate_1	-2.949417	1.47745	-2.00	0.046	-5.84976	-.0490748
_Itime_2	-2.283333	1.683863	-1.36	0.175	-5.588881	1.022215
_cons	23.38	1.381171	16.93	0.000	20.66866	26.09134

- The estimated coefficient on treated is identical to the estimated coefficient in part 6
- The standard error in part 7 is larger, Why?

Part 7

```
. xi: regress emptot treated i.state i.time, robust
i.state           _Istate_0-1         (naturally coded; _Istate_0 omitted)
i.time            _Itime_1-2          (naturally coded; _Itime_1 omitted)

Linear regression                               Number of obs     =        768
                                                F(3, 764)         =       1.35
                                                Prob > F          =     0.2557
                                                R-squared         =     0.0076
                                                Root MSE          =     9.5113
```

emptot	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
treated	2.75	1.842828	1.49	0.136	-.8676072	6.367607
_Istate_1	-2.949417	1.47745	-2.00	0.046	-5.84976	-.0490748
_Itime_2	-2.283333	1.683863	-1.36	0.175	-5.588881	1.022215
_cons	23.38	1.381171	16.93	0.000	20.66866	26.09134

- The estimated coefficient on treated is identical to the estimated coefficient in part 6
- The standard error in part 7 is larger, Why?
- Model in part 6 controls for all variables that vary between restaurants and that are constant over time
  - the variance of the residual is therefore smaller in part 6. then in part 7.

Part 10

```
. regress emptot1 low mid high if state==1 & dna==0, noconstant robust

Linear regression                               Number of obs     =        305
                                                F(3, 302)         =     530.34
                                                Prob > F          =     0.0000
                                                R-squared         =     0.8350
                                                Root MSE          =     9.1328

------------------------------------------------------------------------------
             |               Robust
     emptot1 | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         low |    19.5567   .7746706    25.25   0.000     18.03227    21.08114
         mid |   20.08066   .8423506    23.84   0.000     18.42304    21.73828
        high |      22.25    1.13336    19.63   0.000     20.01972    24.48028
------------------------------------------------------------------------------
```

Part 10

```
. regress emptot1 low mid high if state==1 & dna==0, noconstant robust

Linear regression                               Number of obs     =        305
                                                F(3, 302)         =     530.34
                                                Prob > F          =     0.0000
                                                R-squared         =     0.8350
                                                Root MSE          =     9.1328

------------------------------------------------------------------------------
             |               Robust
     emptot1 | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         low |    19.5567   .7746706    25.25   0.000     18.03227    21.08114
         mid |   20.08066   .8423506    23.84   0.000     18.42304    21.73828
        high |      22.25    1.13336    19.63   0.000     20.01972    24.48028
------------------------------------------------------------------------------
```

- we drop the constant term to include all the dummies
- the estimated coefficients on the dummies give the pre-treatment average employment in restaurants with a low/mid/high wage level

Part 12

```
. regress emptot2 low mid high if state==1 & dna==0, noconstant robust

Linear regression                               Number of obs     =        302
                                                F(3, 299)         =     529.57
                                                Prob > F          =     0.0000
                                                R-squared         =     0.8396
                                                Root MSE          =     9.1215
```

emptot2	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
low	20.87755	1.005764	20.76	0.000	18.89828	22.85682
mid	20.95556	.7565711	27.70	0.000	19.46668	22.44443
high	20.21377	1.022741	19.76	0.000	18.20109	22.22645

- the estimated coefficients on the dummies give the post-treatment average employment in restaurants with a low/mid/high wage level

Part 12

```
. regress emptotd low mid high if state==1 & dna==0, noconstant robust

Linear regression                               Number of obs     =        293
                                                F(3, 290)         =       2.61
                                                Prob > F          =     0.0520
                                                R-squared         =     0.0270
                                                Root MSE          =     7.9965
```

emptotd	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
low	1.204787	.8223238	1.47	0.144	-.4136923	2.823267
mid	.7098485	.6871494	1.03	0.302	-.6425837	2.062281
high	-2.156716	1.004933	-2.15	0.033	-4.134603	-.1788298

Part 12

- the estimated coefficients give the pre-post difference in average employment in restaurants with a low/mid/high wage level
- employment increased in restaurants with an initial low starting wage
- employment decreased in restaurants with an initial high starting wage
- High wage restaurants should not have been affected by the minimum wage (since it is not binding there)
- Hence, we also observe a downward “common” trend here

Part 13

```
. tabulate wage_st wage_st2 if state==1 & dna==0 & high==1
```

WAGE_ST	WAGE_ST2 5.05	5.25	5.28	5.5	5.67	Total
5	38	2	1	3	0	44
5.05	3	1	0	0	0	4
5.06	1	0	0	0	0	1
5.1	1	0	0	0	0	1
5.12	2	1	0	0	0	3
5.15	1	0	0	0	0	1
5.25	5	0	0	1	0	6
5.3	1	0	0	0	0	1
5.42	0	0	0	0	1	1
5.5	5	0	0	1	0	6
5.56	1	0	0	0	0	1
5.62	1	0	0	0	0	1
5.75	0	0	0	1	0	1
Total	59	4	1	6	1	71

- we observe that quite a few high wage restaurants seem to have *decreased* their wages toward the new minimum wage (5.05)
- This questions the validity of the high wage restaurants in NJ as control group
- high wage restaurants also seem to be affected by the change in the minimum wage