IO Assignment Part I

Zikai Xu

October 2023

1. (a) The result of this regression is:

```
Linear regression model:
    investment ~ 1 + HPV_recorded + naics_recode + region + gravity + DAV

Estimated Coefficients:
                     Estimate         SE          tStat       pValue   
                    -----------    ----------    ---------    ----------

    (Intercept)      0.00092586    0.00018759       4.9354    7.9977e-07
    HPV_recorded        0.10482    0.00079297       132.19             0
    naics_recode    -4.0964e-07    2.2596e-05    -0.018129       0.98554
    region           0.00033652    2.1827e-05       15.418     1.252e-53
    gravity         -0.00030805    2.3104e-05      -13.333    1.4903e-40
    DAV                0.089961    0.00020923       429.96             0


Number of observations: 2385887, Error degrees of freedom: 2385881
Root Mean Squared Error: 0.0605
R-squared: 0.0894,  Adjusted R-Squared: 0.0894
F-statistic vs. constant model: 4.68e+04, p-value = 0
```

The estimation tells that the plant's investment decision mainly depends on its HPV status and Depreciated Accumulated Violation. If a plant is recorded as a HPV, it will increase investment probability by 10.48%. And if DAV increase by 1, it will increase investment probability by 9.0%.

(b) The result of this regression using collapsed data is:

```
        Linear regression model:
    compliance ~ 1 + inspection + fine + violation

Estimated Coefficients:
                   Estimate        SE         tStat     pValue
                   ---------    ----------    -------    ------

    (Intercept)      0.96905    0.00013029     7437.5      0   
```

```
    inspection     -0.082591    0.00042398     -194.8      0
    fine            -0.40539     0.0076652    -52.887      0
    violation       -0.87475     0.0016966    -515.59      0


Number of observations: 2385887, Error degrees of freedom: 2385883
Root Mean Squared Error: 0.191
R-squared: 0.126,  Adjusted R-Squared: 0.126
F-statistic vs. constant model: 1.15e+05, p-value = 0
```

My finding is that compliance is decreasing if there is more inspection, fines or violations.

(c) I don't agree that the impact of the strictness of environmental regulations on compliance is negative as the regression shows. It is possible that the regulations are stricter because there are less compliant plants. This endogeneity can result in biases in the estimates.

2. (a) The result of this regression is as follows:

```
Linear regression model:
inspection ~ 1 + DAV*lag_violator_notHPV + DAV*lag_HPV_status +
lag_violator_notHPV*lag_HPV_status

Estimated Coefficients:
                                Estimate        SE         tStat       pValue
                                ---------    ----------    -------    ---------

(Intercept)                     0.08489    0.00019152     443.25          0
DAV                             0.15814     0.0018265     86.585          0
lag_violator_notHPV             0.089317    0.0013233     67.496          0
lag_HPV_status                  0.22481     0.0019321     116.36          0
DAV:lag_HPV_status             -0.043705    0.0024149    -18.098    3.3576e-73


Number of observations: 2355908, Error degrees of freedom: 2355903
Root Mean Squared Error: 0.287
R-squared: 0.0293,  Adjusted R-Squared: 0.0293
F-statistic vs. constant model: 1.78e+04, p-value = 0
```

I find that the probability of inspection on a plant increases significantly if the plant is in HPV status in previous period.

(b) The result of new regression is:

```
Estimated Coefficients:
                            Estimate          SE          tStat       pValue
                            -----------    ----------    -------    ----------

(Intercept)                   0.15112     0.00089575     168.71             0
```

naics_recode	-0.00064205	0.00010782	-5.9548	2.6047e-09
region	-0.010786	0.000104	-103.71	0
gravity	-0.0020765	0.00011005	-18.869	2.0657e-79
DAV	0.15612	0.0018217	85.696	0
lag_violator_notHPV	0.085873	0.0013203	65.038	0
lag_HPV_status	0.22721	0.0019271	117.9	0
DAV:lag_HPV_status	-0.041286	0.0024087	-17.141	7.4191e-66

```
Number of observations: 2355908, Error degrees of freedom: 2355900
Root Mean Squared Error: 0.287
R-squared: 0.0344,  Adjusted R-Squared: 0.0344
F-statistic vs. constant model: 1.2e+04, p-value = 0
```

The main difference is that the intercept is much higher with more controls. The coefficients of lagged regular violator, lagged HPV and DAV are almost the same. This implies that the inspection decision does not depend on these controls.

(c) The regression result is as follows:

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	0.00020174	5.0945e-05	3.96	7.4948e-05
inspection	0.00022399	3.7092e-05	6.0389	1.5522e-09
violation	0.0014815	0.00014828	9.9913	1.6658e-23
naics_recode	-2.46e-05	6.0956e-06	-4.0356	5.4456e-05
region	-1.9356e-05	5.8931e-06	-3.2845	0.0010216
gravity	8.7116e-07	6.2219e-06	0.14002	0.88865
DAV	0.0056033	8.9723e-05	62.451	0
lag_violator_notHPV	-0.00010957	7.4715e-05	-1.4665	0.1425
lag_HPV_status	0.0035842	0.00010936	32.774	1.5818e-235
DAV:lag_violator_notHPV	-0.0040314	0.0001362	-29.6	1.6206e-192

```
Number of observations: 2355908, Error degrees of freedom: 2355898
Root Mean Squared Error: 0.0162
R-squared: 0.00564,  Adjusted R-Squared: 0.00564
F-statistic vs. constant model: 1.49e+03, p-value = 0
```

It turns out that the fine mainly depends on violation, DAV and lagged HPV status. This shows the presence of dynamic enforcement.

3. (a) The regression result is:

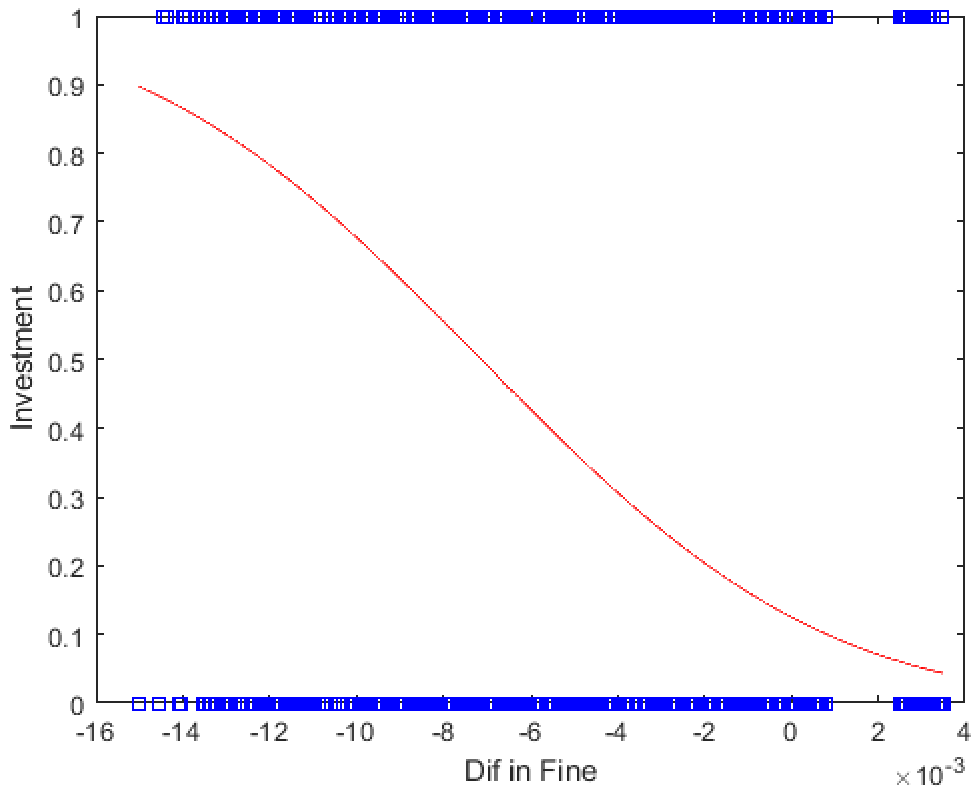
```
Linear regression model:
fine ~ 1 + naics_recode + region + gravity + lag_investment*DAV +
lag_investment*lag_HPV_status + DAV*lag_HPV_status +
lag_investment:DAV:lag_HPV_status

Estimated Coefficients:
                        Estimate          SE          tStat        pValue
                       -----------    ----------    ----------    ---------

(Intercept)             7.946e-05     2.5312e-05       3.1392     0.0016939
lag_investment          4.123e-05     0.00031463      0.13104       0.89574
naics_recode           -6.631e-06     3.0294e-06      -2.1888      0.028608
region                -5.2335e-06      2.941e-06      -1.7795      0.075161
gravity                1.8187e-06     3.1013e-06      0.58644       0.55758
DAV                    0.00022461     0.00012592       1.7837      0.074476
lag_HPV_status         -2.806e-05      0.0063913   -0.0043904        0.9965
lag_investment:DAV    -0.00028857     0.00041611     -0.69348       0.48801
lag_investment:
lag_HPV_status         0.00064911      0.0064024      0.10139       0.91924
DAV:lag_HPV_status    -0.00023072      0.0051121    -0.045132         0.964
lag_investment:
DAV:lag_HPV_status     0.00098014      0.0051317        0.191       0.84853


Number of observations: 2251998, Error degrees of freedom: 2251987
Root Mean Squared Error: 0.00789
R-squared: 3.93e-05,  Adjusted R-Squared: 3.49e-05
F-statistic vs. constant model: 8.86, p-value = 1.03e-14
```

(b) Here is the conditional expected fines:

```
fine_i_v =

    0.0055


fine_i_h =

    0.0090


fine_n_h =

    0.0030
```

```
fine_n_h =

    0.0091
```

I find that the expected fine in the future is higher if the plant is under HPV status. Also the fines are higher if there is no investment. But this is likely because the plant expects that there will be higher fines and then they are more likely to make investment.



(c) The probit model shows that investment will decrease the expected fines in the future.

(d) Here is a simple two period model:
Let $X \in \{0, 1\}$ be the plant's investment decision,
$Fine : \{0, 1\} \to \mathbb{R}_+$ be the expected fines conditional on investment. And let β be the discount factor. Assume that a plant's flow utility is:

$$u(X) = -(\theta^X + \varepsilon)X - Fine(X^-)$$

where X^- is the investment decision in previous period.
So the plant's 2-period problem is:

$$\max_X -(\theta^X + \varepsilon)X - \beta Fine(X)$$

The solution is that $X = 1$ if $-\theta^X - \varepsilon - \beta Fine(1) > -\beta Fine(0)$.
Suppose that ε follows a standard normal distribution, then $P(X = 1) = \Phi(-\beta[Fine(1) - Fine(0)] - \theta^X)$ which is decreasing in $Fine(1) - Fine(0)$ as I showed in part c).

The probit estimate in c) shows that $\hat{\theta}^X = 1.1490$. This estimate is far away from BGL's estimate on investment cost, 2.872. One reason might be that the investment may not only affect the expected fines in next period, but also in further future, hence the 2 period model will underestimate the investment cost.

MATLAB code is attached:

```
%% 1) Reduced-form regressions
% a
% Read data
clc
clear
T = readtable("analysis_data.csv");
%% Run a linear regression
lm = fitlm(T,"investment~region+naics_recode+gravity+HPV_recorded+DAV")

% b
%% Collapse data
tbl = removevars(T,"gravity_str");
varfun(@mean,tbl,'GroupingVariables',{'region','naics_recode','gravity','quarter'});
%% Run a regression using collapsed data
clm = fitlm(tbl,"compliance~inspection+violation+fine")

%% 2) Evidence of dynamic enforcement
% a
ilm = fitlm(T,"inspection~lag_violator_notHPV*lag_HPV_status*DAV")

%% b
blm = fitlm(T,"inspection~lag_violator_notHPV*lag_HPV_status*DAV+region+naics_recode+gravity")

%% c
cclm = fitlm(T,"fine~lag_HPV_status*DAV+region+naics_recode+gravity+inspection+violation")

%% 3)
% a
% Create a table for plants not in compliance
tnic = T(~(T.compliance~=0),:);
% Regression
lm3a = fitlm(tnic,"fine~lag_investment*lag_HPV_status*DAV+region+naics_recode+gravity")

%% b
% Expected fine given investment and regular violator
lag_investment = 1;
naics_recode = 1;
region = 1;
gravity = mean(tnic.gravity);
DAV = mean(tnic.DAV);
lag_HPV_status = 0;
Xnew = table(lag_HPV_status,lag_investment,naics_recode,region,gravity,DAV);
fine_i_v = lm3a.predict(Xnew)
```

```
% Expected fine given investment and HPV
Xnew.lag_HPV_status = 1;
fine_i_h = lm3a.predict(Xnew)

% Expected fine given no investment and regular violator
Xnew.lag_HPV_status = 0;
Xnew.lag_investment = 0;
fine_n_h = lm3a.predict(Xnew)

% Expected fine given no investment and HPV
Xnew.lag_HPV_status = 1;
fine_n_h = lm3a.predict(Xnew)



%% c
X = tnic;
X.lag_investment(:) = 1;
Xinv = X;
X.lag_investment(:) = 0;
Xnoinv = X;

lag_investment = Xinv.lag_investment;
naics_recode = Xinv.naics_recode;
region = Xinv.region;
gravity = Xinv.gravity;
DAV = Xinv.DAV;
lag_HPV_status = Xinv.lag_HPV_status;
Xinv = table(lag_investment,lag_HPV_status,naics_recode,region,gravity,DAV);

lag_investment = Xnoinv.lag_investment;
naics_recode = Xnoinv.naics_recode;
region = Xnoinv.region;
gravity = Xnoinv.gravity;
DAV = Xnoinv.DAV;
lag_HPV_status = Xnoinv.lag_HPV_status;
Xnoinv = table(lag_investment,lag_HPV_status,naics_recode,region,gravity,DAV);
diffine = lm3a.predict(Xinv)-lm3a.predict(Xnoinv);

%% Probit regression
[probitCoef,dev] = glmfit(diffine,tnic.investment,'binomial',"link",'probit')
x = linspace(min(diffine),max(diffine),104446);
probitFit = glmval(probitCoef,x,'probit');
plot(diffine,tnic.investment,'bs', x,probitFit,'r-');
xlabel('Dif in Fine'); ylabel('Investment');
```