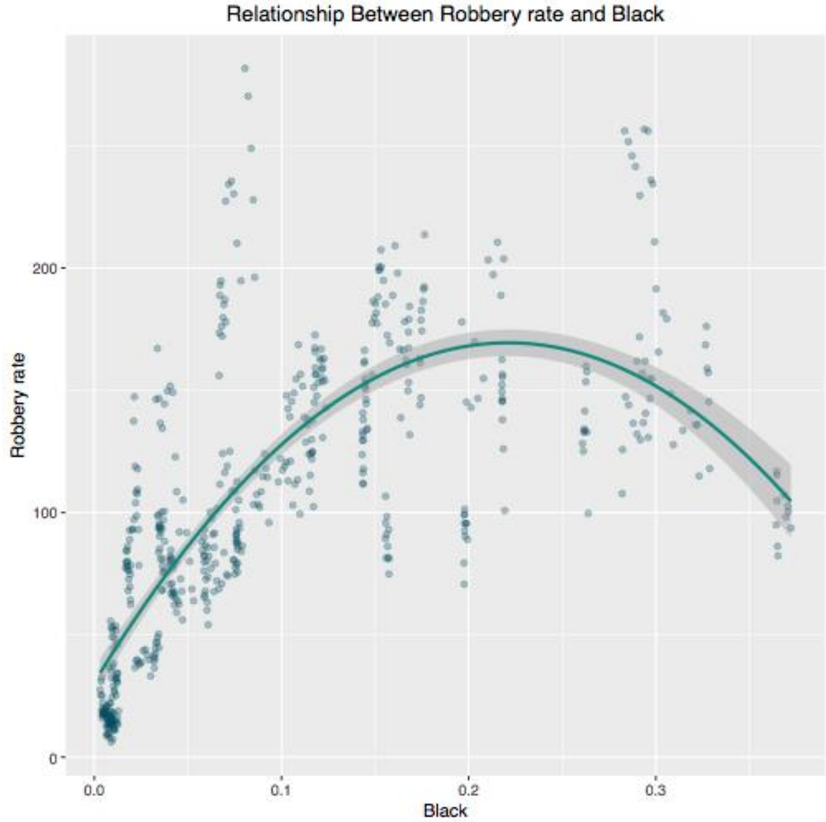
ECOM20001 ASSIGNMENT 3 COVER PAGE

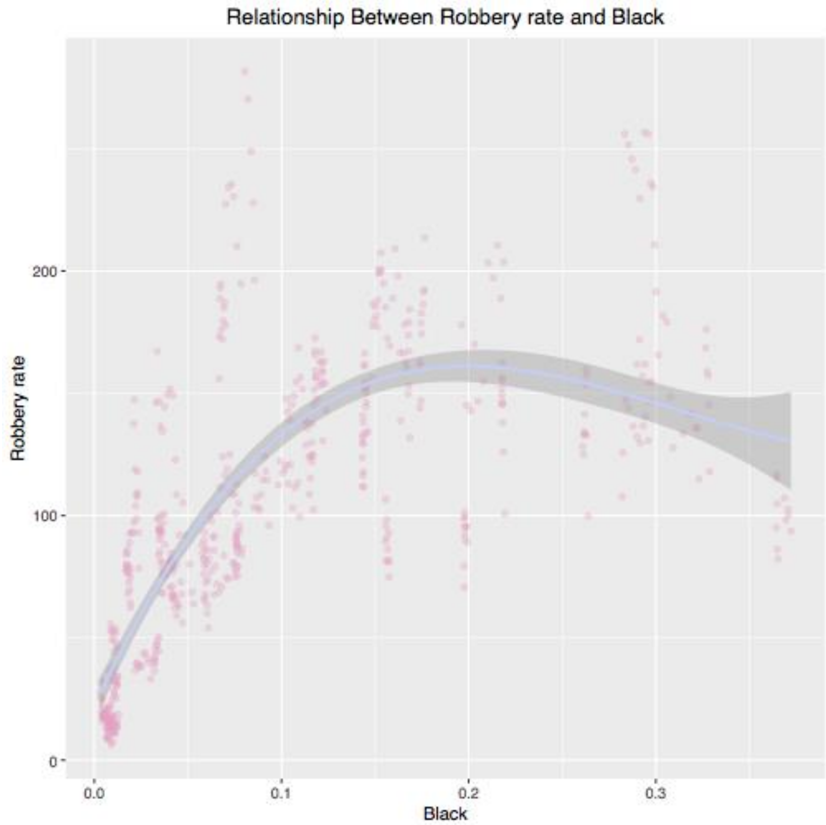
Name	Student ID
Gia Han Ly	1074109
Rui Zheng	1084537
Alexandre Bormann	1072524

Question 1

Quadratic curve fit for relationship between Robbery Rate and Black



Cubic curve fit for relationship between Robbery Rate and Black



There appears to be a non-linear relationship between robbery rate and black variable. We can also visually identify a slight difference between the quadratic and cubic regression line, towards the right end of regression.

Question 2

	Dependent variable:		
	Robbery Rate		
	(1)	(2)	(3)
Black	1,784.90***	1,335.23***	382.61***
	(134.49)	(70.43)	(35.38)
Black Squared	-6,539.92***	-2,916.03***	
	(885.87)	(166.22)	
Black Cube	7,292.21***		
	(1,683.71)		
Income (in 000s)	12.58***	11.89***	19.33***
	(2.63)	(2.57)	(2.46)
Age	-1.54	-2.37*	-2.51
	(1.26)	(1.31)	(2.00)
Female	-1,038.87**	-878.21*	524.14
	(467.54)	(473.92)	(546.97)
d2001	0.99	1.29	1.55
	(7.54)	(7.62)	(9.14)
d2002	-2.20	-1.66	-0.99
	(7.24)	(7.34)	(8.95)
d2003	-5.82	-5.03	-4.89
	(7.23)	(7.30)	(8.92)
d2004	-12.14*	-11.02	-11.11
	(7.08)	(7.14)	(8.88)
d2005	-12.14*	-10.69	-11.65
	(7.17)	(7.23)	(9.25)
d2006	-6.75	-4.98	-6.81
	(7.77)	(7.88)	(9.85)
d2007	-11.31	-9.16	-12.26
	(8.01)	(8.09)	(10.11)
d2008	-13.02	-10.62	-13.63
	(8.10)	(8.15)	(10.21)
d2009	-20.51***	-17.95**	-20.05**
	(7.76)	(7.88)	(9.88)
d2010	-30.65***	-27.88***	-29.74***
	(7.81)	(7.93)	(9.91)
Constant	549.05***	510.07**	-188.65
	(199.10)	(203.25)	(226.58)
Observations	550	550	550
R2	0.67	0.65	0.45
Adjusted R2	0.66	0.64	0.44
Residual Std. Error	34.29 (df = 533)	34.85 (df = 534)	43.73 (df = 535)
F Statistic	66.16*** (df = 16; 533)	67.10*** (df = 15; 534)	31.70*** (df = 14; 535)
Note:			*p<0.1; **p<0.05; ***p<0.01

Run sequential hypothesis testing using t-statistic test in observing whether there is a non-linear relationship between robbery_rate and black.

Let null hypothesis be: coefficient on Black_cu = 0 (there is a linear relationship),

and alternative be coefficient on Black_cu 0 (there is a non-linear relationship)

P-value of coefficient estimate on Black_cu is < 0.0001, therefore reject the null hypothesis at 5% significance level. There exists a non-linear relationship between robbery_rate and black, as we reject the null of linear relationship.

Question 3

a. Increasing black from 0.05 to 0.10 holding all other regressors fixed at: income_scale=1, age=20, female=0.5, d2001=0, d2002=0, d2003=0, d2004=0, d2005=0, d2006=0, d2007=0, d2008=0, d2009=0, d2010=1

> **Let black = 0.05,black_sq=0.0025, black_cu=0.000125**
> Predicted value of Robbery Rate (black=0.05) = 54.6
>
> **Let black = 0.10, black_sq=0.01, black_cu=0.001**
> Predicted value of Robbery Rate (black=0.10) = 101.18
>
> Robbery Rate = 101.18 - 54.6 = 46.58
> F-statistic = 382.62
> SE of partial effect = 46.58/382.62= 2.38
>
> **95% CI for the partial effect on Robbery Rate**
> = [46.58 - 1.96 x 2.38, 46.58 + 1.96 x 2.38]
> = [41.91, 51.24] robbery cases

b. Increasing black from 0.10 to 0.15 holding all other regressors fixed at: income_scale=1, age=20, female=0.5, d2001=0, d2002=0, d2003=0, d2004=0, d2005=0, d2006=0, d2007=0, d2008=0, d2009=0, d2010=1

> **Let black = 0.10, black_sq=0.01, black_cu=0.001**
> Predicted value of Robbery Rate (black=0.10) = 101.18
>
> **Let black = 0.15,black_sq=0.0225, black_cu=0.003375**
> Predicted value of Robbery Rate (black=0.15) = 125.99
>
> Robbery Rate = 125.99 - 101.18 = 24.81
> F-statistics = 162.58
> SE of partial effect = 24.81/162.58 = 1.95
>
> **95% CI for the partial effect on Robbery Rate**
> = [24.81 - 1.96 x 1.95, 24.81 + 1.96 x 1.95)
> = [21.00, 28.63] robbery cases

The partial effect on robbery rate is diminishing as the value of the black variable increases, holding all other variables constant. There exists a positive partial effect on robbery rate with diminishing incremental increases, as the value of black becomes larger. The difference in partial effect is consistent with the non-linear nature of Reg(1), where coefficient estimates on black is positive, indicating a positive partial effect on robbery cases. Furthermore, the negative coefficient on black_sq implies a diminishing partial effect on the regression.

Question 4

	Dependent variable:			
	Log(Robbery Rate)			
	(1)	(2)	(3)	(4)
Share of Pop. that is Black	4.070***			
	(0.432)			
Log of Share of Pop. that is Black		0.590***	0.559***	0.385***
		(0.022)	(0.030)	(0.096)
Log of Share of Pop. that is Black x Years 2004-2007			0.056	
			(0.040)	
Log of Share of Pop. that is Black x Years 2008-2010			0.057	
			(0.043)	
Log of Share of Pop. that is Black x Avg. Household Inc.				0.046**
				(0.020)
Avg. Household Inc. (ten thousands)	0.245***	0.179***	0.183***	0.302***
	(0.031)	(0.023)	(0.023)	(0.056)
Avg. Age	−0.109***	−0.027*	−0.025*	−0.032**
	(0.029)	(0.015)	(0.015)	(0.016)
Share of Pop. that is Female	26.363***	−8.783**	−9.453**	−9.350**
	(6.995)	(4.203)	(4.191)	(4.181)
2001	0.030	−0.001	−0.001	−0.0002
	(0.120)	(0.082)	(0.082)	(0.082)
2002	0.029	−0.031	−0.031	−0.028
	(0.120)	(0.081)	(0.081)	(0.081)
2003	−0.004	−0.078	−0.079	−0.071
	(0.122)	(0.084)	(0.084)	(0.084)
2004	−0.050	−0.148*	0.011	−0.136
	(0.123)	(0.085)	(0.126)	(0.085)
2005	−0.051	−0.163*	−0.005	−0.149*
	(0.126)	(0.084)	(0.129)	(0.083)
2006	0.001	−0.126	0.029	−0.110
	(0.126)	(0.083)	(0.130)	(0.084)
2007	−0.052	−0.186**	−0.032	−0.167*
	(0.131)	(0.088)	(0.130)	(0.088)
2008	−0.048	−0.202**	−0.047	−0.179**
	(0.133)	(0.088)	(0.132)	(0.087)
2009	−0.058	−0.240***	−0.085	−0.219**
	(0.131)	(0.086)	(0.133)	(0.086)
2010	−0.143	−0.344***	−0.190	−0.321***
	(0.134)	(0.087)	(0.135)	(0.087)
Constant	−6.468**	10.850***	10.996***	10.775***
	(2.837)	(1.795)	(1.789)	(1.773)
Observations	550	550	550	550
R2	0.453	0.743	0.745	0.746
Adjusted R2	0.438	0.736	0.737	0.739
Residual Std. Error	0.595 (df = 535)	0.408 (df = 535)	0.407 (df = 533)	0.406 (df = 534)
F Statistic	31.588*** (df = 14; 535)	110.488*** (df = 14; 535)	97.199*** (df = 16; 533)	104.585*** (df = 15; 534)
Note:				*p<0.1; **p<0.05; ***p<0.01

Question 5

Coefficient on black variable

According to coefficient estimate on black in Reg(1), a one unit increase (100% increase) in African-American population, while holding all other independent variables constant, is associated with a 407% percent increase in the number of robbery cases per 100,000 people. Regression coefficient on black is statistically significant at the 5% level, with p-value of less than 0.01.

Coefficient on log_black variable

Number of robberies per 100,000 people increases by 0.590%, for every one percent increase in African-American population, holding all other variables constant. Furthermore, the elasticity of robbery rates with respect to black population is 0.59. Coefficient estimate on log_black in Reg(2) has a p-value < 0.01 therefore, statistically significant at 5% level.

Question 6

Coefficient on log_black_middle variable

Interpreting regression coefficient on log_black_middle in Reg(3), from 2004 to 2007, every one percent increase in the black population is associated with a 0.056% increase in robbery rates per 100,000 people, holding all else constant. Coefficient estimate also suggests an elasticity of 0.056 on robbery rates with respect to black African-American population. The coefficient is statistically insignificant, therefore we fail to reject whether black population in 2004-2007 has no effects on robbery rates.

Coefficient on log_black_end variable

From 2008 to 2010, the coefficient estimates on robbery cases per 100,000 people increased by 0,057% for every 1% increase in African-American population, holding all other constants fixed. The elasticity of black population, during the relevant time period, with respect to robbery rate is 0.057. P-value for the log_black_end variable is greater than 0.05, therefore statistically insignificant at the 5% significance level. We fail to reject that robbery cases are not affected by black population in 2008-2010.

Question 7

Conduct a joint-hypothesis test of the coefficients on log_black_middle and log_black_end. We test joint null that coefficients on both variables are all equal to zero.

F-statistic	0.000308
P-value	0.986
Df 1	1
Df 2	533

Fail to reject the null at 5% significance level, as p-value is greater than 0.05. Therefore, fail to reject whether different years, black_middle (2004-2007) or black_end (2008-2010) have no effect on robbery cases per 100,000 people. Hence, fail to establish whether or not African-American population in different time periods affects robbery cases every 100,000 people.

Question 8

$$\log(\widehat{robbery\,rate}) = \widehat{\beta_0} + \widehat{\beta_1}\log(black) + \widehat{\beta_2}\log(black) * income\_scale + \widehat{\beta_3}\,income\_scale + \widehat{\beta_4}\,age + \widehat{\beta_5}\,female + \widehat{\beta_6}\,d2001 + \widehat{\beta_7}\,d2002 + \widehat{\beta_8}\,d2003 + \widehat{\beta_9}\,d2004 + \widehat{\beta_{10}}\,d2005 + \widehat{\beta_{11}}\,d2006 + \widehat{\beta_{12}}\,d2007 + \widehat{\beta_{13}}\,d2008 + \widehat{\beta_{14}}\,d2009 + \widehat{\beta_{15}}\,d2010$$

Income = $30,000

$$= \widehat{\beta_0} + \widehat{\beta_1}\log(black) + \widehat{\beta_2}\log(black) * 3 + \widehat{\beta_3} * 3 + \widehat{\beta_4}\,age + \widehat{\beta_5}\,female + \widehat{\beta_6}\,d2001 + \widehat{\beta_7}\,d2002 + \widehat{\beta_8}\,d2003 + \widehat{\beta_9}\,d2004 + \widehat{\beta_{10}}\,d2005 + \widehat{\beta_{11}}\,d2006 + \widehat{\beta_{12}}\,d2007 + \widehat{\beta_{13}}\,d2008 + \widehat{\beta_{14}}\,d2009 + \widehat{\beta_{15}}\,d2010$$

$$= \widehat{\beta_0} + (\widehat{\beta_1} + 3\widehat{\beta_2}) * \log(black) + 3\widehat{\beta_3} + \widehat{\beta_4}\,age + \widehat{\beta_5}\,female + \widehat{\beta_6}\,d2001 + \widehat{\beta_7}\,d2002 + \widehat{\beta_8}\,d2003 + \widehat{\beta_9}\,d2004 + \widehat{\beta_{10}}\,d2005 + \widehat{\beta_{11}}\,d2006 + \widehat{\beta_{12}}\,d2007 + \widehat{\beta_{13}}\,d2008 + \widehat{\beta_{14}}\,d2009 + \widehat{\beta_{15}}\,d2010$$

<u>Elasticity</u>

$$(\widehat{\beta_1} + 3\widehat{\beta_2}) = 0.3853 + 3 * 0.0464$$
$$= 0.52466$$

<u>Testing for elasticity</u>

$$H_0: \widehat{\beta_1} + 3\widehat{\beta_2} = 0; \quad H_1: \widehat{\beta_1} + 3\widehat{\beta_2} \neq 0$$

$$\text{SE} = \frac{0.52466}{\sqrt{184.0905}} = 0.03867$$

F-stat = 184.09
p-value < 0.05
Df 1 = 1, Df 2 = 534

95% CI for elasticity of Robbery Rate with respect to Black given income = $30,000
[0.52466 - 1.96 x 0.03867, 0.52466 + 1.96 x 0.03867]
[0.449, 0.600]

The elasticity of robbery rate with respect to black, where income = $30,000, is the sum of the coefficients on log(black). Based on the estimated coefficients in Reg(4), the elasticity/partial effect of the relationship between robbery rate and black is 0.525. Every one percent increase in black is associated with a 0.525% increase in robbery rates, given income =$30,000 and all else held constant.

Income = $50,000

$$= \widehat{\beta_0} + \widehat{\beta_1}\log(black) + \widehat{\beta_2}\log(black) * 5 + \widehat{\beta_3} * 5 + \widehat{\beta_4}\, age + \widehat{\beta_5}\, female$$
$$+ \widehat{\beta_6}\, d2001 + \widehat{\beta_7}\, d2002 + \widehat{\beta_8}\, d2003 + \widehat{\beta_9}\, d2004 + \widehat{\beta_{10}}\, d2005 + \widehat{\beta_{11}}\, d2006$$
$$+ \widehat{\beta_{12}}\, d2007 + \widehat{\beta_{13}}\, d2008 + \widehat{\beta_{14}}\, d2009 + \widehat{\beta_{15}}\, d2010$$
$$= \widehat{\beta_0} + \left(\widehat{\beta_1} + 5\widehat{\beta_2}\right) * \log(black) + 5\widehat{\beta_3} + \widehat{\beta_4}\, age + \widehat{\beta_5}\, female + \widehat{\beta_6}\, d2001$$
$$+ \widehat{\beta_7}\, d2002 + \widehat{\beta_8}\, d2003 + \widehat{\beta_9}\, d2004 + \widehat{\beta_{10}}\, d2005 + \widehat{\beta_{11}}\, d2006 + \widehat{\beta_{12}}\, d2007$$
$$+ \widehat{\beta_{13}}\, d2008 + \widehat{\beta_{14}}\, d2009 + \widehat{\beta_{15}}\, d2010$$

Elasticity

$$\left(\widehat{\beta_1} + 5\widehat{\beta_2}\right) = 0.3853 + 5 * \ 0.0464$$
$$= 0.61726$$

Testing for elasticity
$H_0\colon \widehat{\beta_1} + 5\widehat{\beta_2} = 0; \quad H_1\colon \widehat{\beta_1} + 5\widehat{\beta_2} \neq 0$

SE = $\frac{0.61726}{\sqrt{715.4346}}$ = 0.02308

F-stat = 715.43
p-value < 0.05
Df 1 = 1, Df 2 = 534

95% CI for elasticity of Robbery Rate with respect to Black given income = $50,000
[0.61726 - 1.96 x 0.02308, 0.61726 + 1.96 x 0.02308]
[0.572, 0.662]

Similarly to where income = $30,000, the sum of log(black) coefficient implies the elasticity of robbery rates with respect to variable black. The elasticity, calculated with the estimated coefficients from Reg(4), is 0.617. This implies that robbery rate increases by 0.617% for every one percent increase in African-American population, given that income = $50,000 and all else is held constant.

Proposal of Method of Relaxing Social Distancing Law

The following investigation will outline a method for testing the impact of relaxing COVID-19 related restrictions in Australia. More specifically, we will be focusing on allowing up to 10 dine-in customers in restaurants of the treatment Local Government Areas (LGAs). The timeline considered for the investigation will begin on the first of May. The investigation would be in alignment with the “three phases”, Loosening of Restriction proposal, the Australian Government has been working on. This will yield a method that can directly be applied across different states.

Research Design

Currently, there are several states, such as New South Wales and Victoria, who are lifting their social distancing restrictions at different rates. We will try to mimic the process of the government and allow for a two-week observation period before the treatment group will be allowed to implement dine-in customers. For our investigation, we will compare the change in infection rate across 100 isolated LGAs in Australia. This will allow us to maintain IID sampling, preventing contamination between treatment groups and the control groups.

Another crucial aspect of our research design is that the investigation will remain unbiased. This allows us to use our collected sample data, to estimate the population values, for the effect of loosening restrictions across Australia. We will be randomly selecting LGAs with a population between 100,000 and 200,000 people. As a result, it would ensure a selection of LGAs with similar population density, infrastructure and infection rate. From doing so, it’ll allow us to conduct a difference in differences test, to determine the impact of relaxing COVID-19 restrictions on the spread of the virus, under identical conditions and characteristics across LGAs.

Data Collection

There are several factors to consider when collecting data for our research. These include: the level of aggregation, the frequency, and the sample period, for collecting the data. For our aggregation level, we decided to go with measuring cases by LGAs. 100 LGAs will be randomly selected, and each LGAs will be randomly assigned to either the treatment group, which will allow up to 10 dine-in customers, or the control group, which will not be allowed to eat in restaurants at all. This process will be done using a random number generator, to ensure that the data remains unbiased. There will then be 50 LGAs in the treatment group and 50 in the control group.

Our sample period for collecting the data will be one month. This is in line with the timeline the government has followed to relax COVID-19 related restrictions. We will be assuming that our data was collected in May, as that is when the first state transferred from allowing no customers in restaurants, to allowing some. During the first two weeks, both treatment and control groups will be subjected to maximum restriction, with takeaway as the only viable dining option. After two weeks, the restaurants in the treatment group will then be allowed to host up to 10 guests at a time.

In allowing the treatment group to be only exposed for two weeks, we are trying to follow the same steps as some states, who are relaxing their social distancing restrictions with a two-week difference to other states, to gradually ease covid-19 restrictions across Australia. This also enabled other states to observe the changes in COVID-19 cases, and also be able to conduct their own difference in differences test, to a state who has lifted their restrictions weeks prior.

The frequency of the COVID-19 test will be every day. The people in each LGA will be selected randomly, and the number of tests conducted will be limited to the number of test kits that are available. Although the data will be collected daily, the results will only be interpreted weekly. This is to account for the high transmission rate of COVID-19 and also the delay in symptoms. There might be large spikes in the number of COVID-19 cases some day of the week, and to aggregate those spikes and eliminate possible outliers, it is more reasonable to record the case numbers after every week.

Model

$$log\ (y_{it}) = \beta_0 + \beta_1 Treatment_{it} + \beta_2 Post_{it} + \beta_3 (Treatment_{it} \times Post_{it}) + u_{it}$$

Where:

- y is the infection rate
- β_0 is the intercept for the control group before relaxing the restrictions
- β_1 is the difference between treatment and control group initially
- β_2 is the difference in the control group before and after relaxing the restrictions
- β_3 is the difference in difference between the control and treatment group after relaxing the restrictions
- Treatment is a dummy variable where treatment = 1 if the restriction is relaxed and treatment = 0 if not
- Post is a dummy variable where post = 1 if the time period is after the restriction is relaxed and post = 0 if before

Our proposed regression is a log-linear model, this allows us to compare the percentage change in the number of cases, based on whether the restrictions, of not allowing people to dine in restaurants, were lifted. We can take the difference in differences test to estimate the effect of lifting restrictions on the number of cases across LGAs in the sample. The coefficient of interest for quantifying that change is the interaction term β_3.

In order to interpret the model, once the data has been collected, we will be able to estimate the effect of allowing 10 diners in restaurants of the treatment group. The change will then be quantifiable by looking how case number increases when people have been allowed to eat inside restaurants β_3. We will then be able to estimate the % change in total infections based on the number obtained for the interaction term.

Our model does not take into account controls (covariates) due to the fact that the LGAs have been randomly assigned to either the treatment or control group. Due to random sampling, the risk factors for both groups are assumed to be similar. Furthermore, the control group itself has the same covariates as the treatment group except for its new policy of dining in restaurants. Controls could however increase the precision of our model by reducing variation in the risk factors. Our model could also face problems of external validity such as age, income or a dummy variable for health insurance and other trends not observed by the model. This might limit the extent to which the results can be applied to other towns; as well as either overstated or understated the difference in difference results that the model produced.

Appendix

#Assignment 3
#***
#Load R Packages
```
library(AER)
library(stargazer)
library(ggplot2)
```

#load dataset
```
mydata1=read.csv(file="as3_crime.csv")
```

#Create new variables
```
mydata1$rob2=mydata1$robbery_rate*mydata1$robbery_rate
mydata1$rob3=mydata1$robbery_rate*mydata1$robbery_rate*mydata1$robbery_rate
```

#Question 1
#fitting quadratic function onto scatterplot using ggplot
```
pdf("meo2.pdf")
ggplot(mydata1,aes(y=robbery_rate, x=black)) +
  geom_point(alpha = .3, color = "#014d64") +
  stat_smooth(method = "lm", formula = y ~ poly(x,2), color = "#00887d") +
  ggtitle("Relationship Between Robbery rate and Black") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_x_continuous(name="Black") +
  scale_y_continuous(name="Robbery rate")
dev.off()
#fitting cubic function onto scatterplot using ggplot
pdf("meo3.pdf")
ggplot(mydata1,aes(y=robbery_rate, x=black)) +
  geom_point(alpha = .3, color = "#E6A0C4") +
  stat_smooth(method = "lm", formula = y ~ poly(x,3), color = "#C6CDF7") +
  ggtitle("Relationship Between Robbery rate and Black") +
  theme(plot.title = element_text(hjust = 0.5)) +
  scale_x_continuous(name="Black") +
  scale_y_continuous(name="Robbery rate")
dev.off()
```

#Question 2
```
mydata1$income_scale=mydata1$income/10000
mydata1$black_sq=mydata1$black*mydata1$black
mydata1$black_cu=mydata1$black*mydata1$black*mydata1$black
mydata1$d2000=as.numeric(mydata1$year==2000)
mydata1$d2001=as.numeric(mydata1$year==2001)
mydata1$d2002=as.numeric(mydata1$year==2002)
mydata1$d2003=as.numeric(mydata1$year==2003)
mydata1$d2004=as.numeric(mydata1$year==2004)
mydata1$d2005=as.numeric(mydata1$year==2005)
mydata1$d2006=as.numeric(mydata1$year==2006)
mydata1$d2007=as.numeric(mydata1$year==2007)
mydata1$d2008=as.numeric(mydata1$year==2008)
mydata1$d2009=as.numeric(mydata1$year==2009)
mydata1$d2010=as.numeric(mydata1$year==2010)
```

```
#Robbery_rate regressions with cubic functions
reg1=lm(robbery_rate~black+black_sq+black_cu+income_scale+age+female+d2001+
      d2002+d2003+d2004+d2005+d2006+d2007+d2008+d2009+d2010,data=mydata1)
cov1=vcovHC(reg1, type = "HC1")
se1=sqrt(diag(cov1))
#Robbery_rate regressions with quadratic functions
reg2=lm(robbery_rate~black+black_sq+income_scale+age+female+d2001+
       d2002+d2003+d2004+d2005+d2006+d2007+d2008+d2009+d2010,data=mydata1)
cov2=vcovHC(reg2, type = "HC1")
se2=sqrt(diag(cov2))
#Robbery_rate regressions with linear functions
reg3=lm(robbery_rate~black+income_scale+age+female+d2001+
       d2002+d2003+d2004+d2005+d2006+d2007+d2008+d2009+d2010,data=mydata1)
cov3=vcovHC(reg3, type = "HC1")
se3=sqrt(diag(cov3))
#regression output table for the 3 regressions
stargazer(reg1, reg2, reg3, type="text",
       se=list(se1, se2, se3),
       digits=2,
       dep.var.labels=c("Robbery Rate"),
       covariate.labels=
        c("Black",
         "Black Squared",
         "Black Cube",
         "Income (in 000s)",
         "Age",
         "Female",
         "d2001",
         "d2002",
         "d2003",
         "d2004",
         "d2005",
         "d2006",
         "d2007",
         "d2008",
         "d2009",
         "d2010",
         "Constant"),
       out="reg_q2_output.txt")
#run t-stat test to prove non-linear relationship
 t.test(mydata1$black_cu)
```

#Question 3

```
#part a
#construct data frame for predicting robbery rate given black = 0.05
newdata1=data.frame(black=0.05, black_sq=0.0025, black_cu=0.000125,
            income_scale=1, age=20, female=0.5,
            d2001=0, d2002=0, d2003=0, d2004=0, d2005=0,
            d2006=0, d2007=0, d2008=0, d2009=0, d2010=1)
#construct data frame for predicting robbery rate given black = 0.10
newdata2=data.frame(black=0.10, black_sq=0.01, black_cu=0.001,
```

```
                 income_scale=1, age=20, female=0.5,
                 d2001=0, d2002=0, d2003=0, d2004=0, d2005=0,
                 d2006=0, d2007=0, d2008=0, d2009=0, d2010=1)
#compute predicted value given black = 0.05 and 0.10
robb1=predict(reg1, newdata = newdata1)
robb2=predict(reg1, newdata = newdata2)
#compute partial effect on robbery rate for changes on black from 0.05 to 0.10
drobb=robb2-robb1
#Fstatistic for the joint test of the null
Ftest=linearHypothesis(reg1,c("0.05*black+0.0075*black_sq+0.000875*black_cu=0")
                   ,vcov = v-covHC(reg1, "HC1"))
Fstat=Ftest[2,3]
se_drobb=abs(drobb)/sqrt(Fstat)
#constructing 95% CI for partial effect
drobb_ci95L=drobb-se_drobb*1.96
drobb_ci95H=drobb+se_drobb*1.96
drobb_ciwidth=drobb_ci95H-drobb_ci95L
#Outputting results
sprintf("partial effect: %f", drobb)
sprintf("SE of partial effect: %f", se_drobb)
sprintf("95 CI lower bound for partial effect: %f", drobb_ci95L)
sprintf("95 CI upper bound for partial effect: %f", drobb_ci95H)
sprintf("Width of 95 CI for partial effect: %f", drobb_ciwidth)

#part b
#construct data frame for predicting robbery rate given black = 0.15
newdata3=data.frame(black=0.15, black_sq=0.0225, black_cu=0.003375,
                 income_scale=1, age=20, female=0.5,
                 d2001=0, d2002=0, d2003=0, d2004=0, d2005=0,
                 d2006=0, d2007=0, d2008=0, d2009=0, d2010=1)
#compute predicted value given black = 0.15
robb3=predict(reg1, newdata = newdata3)
#compute partial effect on robbery rate for changes on black from 0.10 to 0.15
drobb1=robb3-robb2
#F-statistic for the joint test of the null
Ftest1=linearHypothesis(reg1,c("0.05*black+0.0125*black_sq+0.002375*black_cu=0"),
                    vcov = vcovHC(reg1, "HC1"))
Fstat1=Ftest1[2,3]
se_drobb1=abs(drobb1)/sqrt(Fstat1)
#constructing 95% CI for partial effect
drobb1_ci95L=drobb1-se_drobb1*1.96
drobb1_ci95H=drobb1+se_drobb1*1.96
drobb1_ciwidth=drobb1_ci95H-drobb1_ci95L

# Outputting results
sprintf("partial effect: %f", drobb1)
sprintf("SE of partial effect: %f", se_drobb1)
sprintf("95 CI lower bound for partial effect: %f", drobb1_ci95L)
sprintf("95 CI upper bound for partial effect: %f", drobb1_ci95H)
sprintf("Width of 95 CI for partial effect: %f", drobb1_ciwidth)
```

#Question 4

```
#Logarithmic variables
mydata1$log_robbery_rate = log(mydata1$robbery_rate)
mydata1$log_black = log(mydata1$black)
#Dummy variables
mydata1$start=1*(mydata1$year<=2003)
mydata1$middle=1*(mydata1$year>=2004&mydata1$year<=2007)
mydata1$end=1*(mydata1$year>=2008)
#Interactions variables
mydata1$log_black_income = mydata1$log_black*mydata1$income_scale
mydata1$log_black_start = mydata1$log_black*mydata1$start
mydata1$log_black_middle = mydata1$log_black*mydata1$middle
mydata1$log_black_end = mydata1$log_black*mydata1$end
#Log-Linear regression
reg1.4=lm(log_robbery_rate~black+income_scale+age+female+d2001+d2002+d2003+d2004+
     d2005+d2006+d2007+d2008+d2009+d2010, data=mydata1)
cov1.4=vcovHC(reg1.4, type = "HC1")
se1.4=sqrt(diag(cov1.4))
#Log-Log regression
reg2.4=lm(log_robbery_rate~log_black+income_scale+age+female+d2001+d2002+d2003+d2004+
      d2005+d2006+d2007+d2008+d2009+d2010, data=mydata1)
cov2.4=vcovHC(reg2.4, type = "HC1")
se2.4=sqrt(diag(cov2.4))
#Log-Log regression with black_middle and black_end
reg3.4=lm(log_robbery_rate~log_black+log_black_middle+log_black_end+income_scale+age+
      female+d2001+d2002+d2003+d2004+d2005+d2006+d2007+d2008+d2009+
      d2010, data=mydata1)
cov3.4=vcovHC(reg3.4, type = "HC1")
se3.4=sqrt(diag(cov3.4))
#Log-Log regression with bla
reg4.4=lm(log_robbery_rate~log_black+log_black_income+income_scale+age+female+d2001+
     d2002+d2003+d2004+d2005+d2006+d2007+d2008+d2009+d2010, data=mydata1)
cov4.4=vcovHC(reg4.4, type = "HC1")
se4.4=sqrt(diag(cov4.4))
#Polynomial regression output table
stargazer(reg1.4, reg2.4, reg3.4, reg4.4, type="text",
     se=list(se1.4, se2.4, se3.4, se4.4),
     align=TRUE, dep.var.labels=c("Log(Robbery Rate)"),
     covariate.labels=
      c("Share of Pop. that is Black",
       "Log of Share of Pop. that is Black",
       "Log of Share of Pop. that is Black x Years 2004-2007",
       "Log of Share of Pop. that is Black x Years 2008-2010",
       "Log of Share of Pop. that is Black x Avg. Household Inc.",
       "Avg. Household Inc. (ten thousands)",
       "Avg. Age",
       "Share of Pop. that is Female",
       "2001",
       "2002",
       "2003",
       "2004",
       "2005",
       "2006",
```

```
          "2007",
          "2008",
          "2009",
          "2010",
          "Constant"),
        out="q4_reg_output.txt")

#Question 7
linearHypothesis(reg3.4,c("log_black_middle=log_black_end"), vcov = vcovHC (reg3.4, "HC1"))

#Question 8
#B0 + B1 log(black) + B2 log(black)*income_scale
summary8=summary(reg4.4)
summarycf=summary8[["coefficients"]]
#Elasticity at income=3
b1=summarycf[2,1]
b2=summarycf[3,1]
E3=b1+3*b2
#F-statistic for the joint test of the null
Ftest8.1=linearHypothesis(reg4.4, c("log_black+3*log_black_income=0"),
                  vcov = vcovHC (reg4.4, "HC1"))
Fstat8.1=Ftest8.1[2,3]
se_e3=E3/sqrt(Fstat8.1)
#Compute 95% CI for elasticity of robbery rates
E3_95ci_l=E3-1.96*se_e3
E3_95ci_h=E3+1.96*se_e3
E3_95ci_w=E3_95ci_h-E3_95ci_l
#Result
sprintf("elasticity at income = 3: %f", E3)
sprintf("SE of elasticity at income = 3: %f", se_e3)
sprintf("95 CI lower bound E3: %f", E3_95ci_l)
sprintf("95 CI upper bound for E3: %f", E3_95ci_h)
sprintf("Width of 95 CI for E3: %f", E3_95ci_w)
#Elasticity at income=5
E5=b1+5*b2
#statistic for the joint test of the null
Ftest8.2=linearHypothesis(reg4.4, c("log_black+5*log_black_income=0"),
                  vcov = vcovHC (reg4.4, "HC1"))
Fstat8.2=Ftest8.2[2,3]
se_e5=E5/sqrt(Fstat8.2)
#Compute 95% CI for elasticity of robbery rates
E5_95ci_l=E5-1.96*se_e5
E5_95ci_h=E5+1.96*se_e5
E5_95ci_w=E5_95ci_h-E5_95ci_l
#Result
sprintf("elasticity at income = 5: %f", E5)
sprintf("SE of elasticity at income = 5: %f", se_e5)
sprintf("95 CI lower bound E5: %f", E5_95ci_l)
sprintf("95 CI upper bound for E5: %f", E5_95ci_h)
sprintf("Width of 95 CI for E5: %f", E5_95ci_w)
```