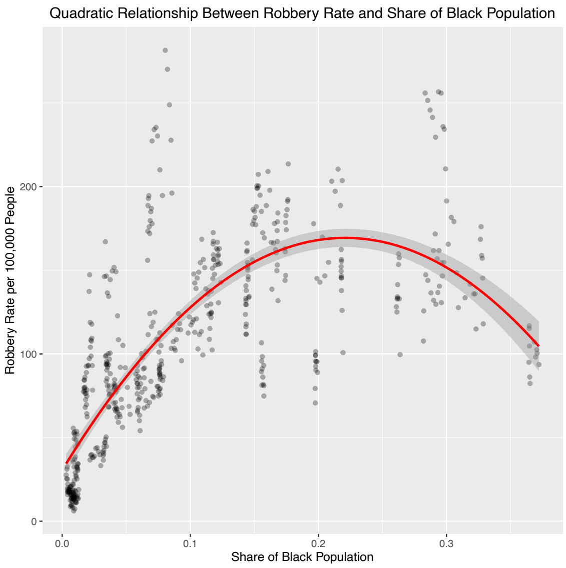
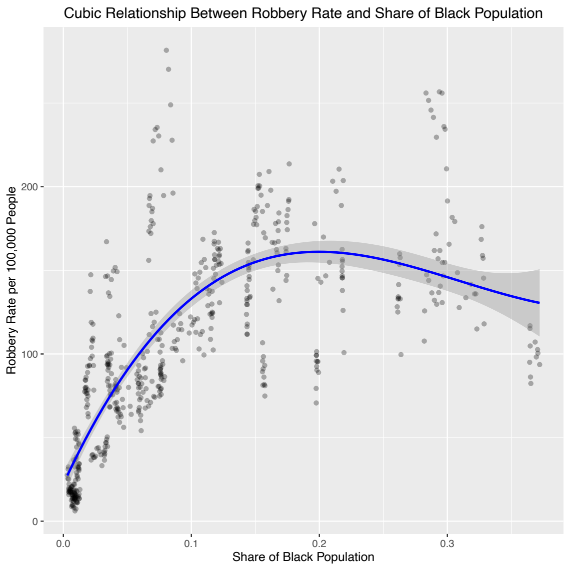
ECOM20001: Econometrics 1

Assignment 3: Suggested Solutions

1. Plots provided below. We can see that the quadratic line of best fit implies a sharper turn in the parabola implying a decreasing relationship between robbery_rate and black after around the point where black = . The cubic allows for more flexibility in the relationship and yields a less sharp parabola and is more in-line with an increasing relationship that grows quickly and then diminishes as the value for black becomes large.





2. Regression results from stargazer() produced on the next page. The more straightforward way to see if there is a nonlinear relationship between robbery_rate and black is to see whether the coefficient on black_sq in Reg(2) is statistically significantly different from 0. In column (2) of the Table, we see that the coefficient estimate of -2,916.029 is indeed statistically significantly different from 0 at the 5% level implying, statistically, that a non-linear relationship exists.[1]

[1] We would likewise conclude, statistically, a non-linear relationship exists based on the column (1) coefficients on black_sq and black_cu. It is perfectly fine to draw this conclusion based on these estimates as well either via individual tests on either coefficient, or a joint test based on both coefficients jointly equalling 0.

Question 2 Regression Results: Polynomial Regression

	Dependent variable:		
	Robbery Rate		
	(1)	(2)	(3)
Share of Pop. that is Black	1,784.902***	1,335.229***	382.614***
	(134.491)	(70.427)	(35.377)
Share of Pop. that is Black Squared	-6,539.925***	-2,916.029***	
	(885.871)	(166.221)	
Share of Pop. that is Black Cubed	7,292.209***		
	(1,683.707)		
Avg. Household Inc. (ten thousands)	12.576***	11.889***	19.335***
	(2.631)	(2.569)	(2.456)
Avg. Age	-1.537	-2.371*	-2.505
	(1.260)	(1.313)	(2.003)
Share of Pop. that is Female	-1,038.867**	-878.206*	524.144
	(467.541)	(473.915)	(546.971)
2001	0.990	1.290	1.546
	(7.539)	(7.620)	(9.142)
2002	-2.198	-1.662	-0.994
	(7.236)	(7.341)	(8.955)
2003	-5.824	-5.029	-4.891
	(7.232)	(7.299)	(8.918)
2004	-12.139*	-11.024	-11.105
	(7.075)	(7.140)	(8.878)
2005	-12.143*	-10.688	-11.650
	(7.173)	(7.234)	(9.252)
2006	-6.752	-4.978	-6.812
	(7.773)	(7.882)	(9.853)
2007	-11.309	-9.160	-12.263
	(8.015)	(8.094)	(10.109)
2008	-13.019	-10.616	-13.625
	(8.101)	(8.149)	(10.207)
2009	-20.509***	-17.949**	-20.052**
	(7.760)	(7.881)	(9.879)
2010	-30.649***	-27.879***	-29.742***
	(7.812)	(7.933)	(9.907)
Constant	549.051***	510.066**	-188.652
	(199.096)	(203.255)	(226.578)
Observations	550	550	550
R2	0.665	0.653	0.453
Adjusted R2	0.655	0.644	0.439
Residual Std. Error	34.291 (df = 533)	34.854 (df = 534)	43.725 (df = 535)
F Statistic	66.155*** (df = 16; 533)	67.100*** (df = 15; 534)	31.702*** (df = 14; 535)

3. In changing black from 0.05 to 0.10, we follow the approach from the first part of lecture note 8 (slides 23, 24 summarise them).

   - The change in robbery_rate from changing black from 0.05 to 0.10 predicted by the cubic regression equation in Reg (1) is computed as **dY**=(1784.902 x 0.10 +6539.925 x 0.10 x 0.10-7292.209 x 0.10 x 0.10 x 0.10) - (1784.902 x 0.05 +6539.925 x 0.05 x 0.05-7292.209 x 0.05 x 0.05 x 0.05)=46.576.

   - Now we compute the standard error of the predicted effect. Letting b1 be the coefficient on black, b2 be the coefficient on black_sq, and b3 be the coefficient on black_cu the general formula for the partial effect **dY** from changing black from 0.05 to 0.10 is **dY=**(b1 x 0.10 + b2 x 0.10 x 0.10 + b3 x 0.10 x 0.10 x 0.10)-(b1x 0.05 + b2 x 0.05 x 0.05+ b3 x 0.05 x 0.05 x 0.05)=0.05 x b1 + 0.0075 x b2 + 0.000875 x b3. Simplifying this expression, with Reg(1) we test the joint null that 0.05 x b1 + 0.0075 x b2 + 0.000875 x b3 = 0. We obtain an F-statistic of **F**=382.6196 with 1 and 533 df. This implies a standard error of the predicted effect of SE(**dY**)= abs(**dY**)/sqrt(**F**)=abs(46.576)/sqrt(382.6196)=2.381

   - 95% CI = [46.576-1.96*2.381,46.576+1.96*2.381]=[41.909, 51.243]. Summarising the results in words: the change in black from 2 to 4 increases the annual robbery_rate by 47 robberies per 100,000 people, with a 95%CI [42, 51] robberies per 100,000 people.

We can compute the partial effect of changing black from 0.10 to 0.15 using the exact same steps. See as3.R for details on the calculations. We obtain a partial effect of a 24.815 increase in robbery_rate, with a standard error of 1.946 and a 95% CI of [21.000, 28.630]

In summary, we find that after controlling for other factors, unlike what the figures suggest in question 1, we in fact find that the robbery rate is <u>decreasing</u> with the share of the population that is black when black goes from 0.05 to 0.10 compared to when it goes from 0.10 to 0.15. This is in-line with the ggplot() graphs above.

4. Regression results from stargazer() produced below.

	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*** (0.022)	0.559*** (0.030)	0.385*** (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*** (0.031)	0.179*** (0.023)	0.183*** (0.023)	0.302*** (0.056)
Avg. Age	−0.109*** (0.029)	−0.027* (0.015)	−0.025* (0.015)	−0.032** (0.016)
Share of Pop. that is Female	26.363*** (6.995)	−8.783** (4.203)	−9.453** (4.191)	−9.350** (4.181)
2001	0.030 (0.120)	−0.001 (0.082)	−0.001 (0.082)	−0.0002 (0.082)
2002	0.029 (0.120)	−0.031 (0.081)	−0.031 (0.081)	−0.028 (0.081)
2003	−0.004 (0.122)	−0.078 (0.084)	−0.079 (0.084)	−0.071 (0.084)
2004	−0.050 (0.123)	−0.148* (0.085)	0.011 (0.126)	−0.136 (0.085)
2005	−0.051 (0.126)	−0.163* (0.084)	−0.005 (0.129)	−0.149* (0.083)
2006	0.001 (0.126)	−0.126 (0.083)	0.029 (0.130)	−0.110 (0.084)
2007	−0.052 (0.131)	−0.186** (0.088)	−0.032 (0.130)	−0.167* (0.088)
2008	−0.048 (0.133)	−0.202** (0.088)	−0.047 (0.132)	−0.179** (0.087)
2009	−0.058 (0.131)	−0.240*** (0.086)	−0.085 (0.133)	−0.219** (0.086)
2010	−0.143 (0.134)	−0.344*** (0.087)	−0.190 (0.135)	−0.321*** (0.087)
Constant	−6.468** (2.837)	10.850*** (1.795)	10.996*** (1.789)	10.775*** (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 4 Regression Results: Logarithmic and Interactive Regressions

5. The Reg(1) coefficient from a log-linear model implies a one-unit change in black yields a corresponding 407% increase in robbery_rate![2] The Reg (2) coefficient from a log-log model implies that a 1% increase in yields a corresponding 0.59% increase in robbery_rate. From the table, both of these respective

[2] This is indeed the literal interpretation from the regression, and it corresponds to a relatively non-sensical one-unit change in which corresponds to a state going from 0% black citizens to 100% black citizens. It highlights in part that the log-linear specification is not as useful in this setting with these regressions from an interpretability standpoint.

regression coefficients have p-values less than 0.01 implying that they are both statistically significantly different from 0.

6. Given the regression set-up, the coefficient on log_black corresponds to the elasticity of robbery_rate with respect to black for years where start=1. This means then that the individual coefficients on log_black_middle and log_black_end in represent the incremental change in this elasticity in years where middle=1 and where black=1 above and beyond the estimate baseline elasticity on in the black regression. From the table, the individual coefficient estimates on log_black_middle and log_black_end in Reg(3) are 0.056 and 0.057, respectively. Quantitatively, they imply that the elasticity of robbery_rate with respect to black is 0.556+0.056=0.612 in the middle (2004-2007) and 0.556+0.057=0.613 end (2008-2010) years of the sample. However, both the individual coefficient estimates in Reg(3) on log_black_middle and log_black_end in Reg(3) are individually statistically insignificantly different from 0 as per the regression output above. In plain language, this means that the elasticity of robbery_rate with respect to black is unchanged in the middle and end years of the sample relative to what we find in the start (2000-2003) years of the sample.

7. We obtain a F-statistic for the test of 3e-04 or 0.00003 with a p-value 0.986 which implies we fail to reject the null that the coefficients on log_black_middle and log_black_end in Reg(3) are equal. The F-statistic is distributed as F(q,n-k-1), and with n=550 observations, k=16 regressors and q=1 restrictions, this implies a F(1,550-16-1) = F(1,533) distribution. In words, the elasticity of robbery_rate with respect to black is unchanged between middle (2004-2007) and end (2008-2010) years of the sample.

```
> linearHypothesis(reg3,c("log_black_middle=log_black_end"),vcov = vcovHC(reg3, "HC1"))
Linear hypothesis test

Hypothesis:
log_black_middle - log_black_end = 0

Model 1: restricted model
Model 2: 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

Note: Coefficient covariance matrix supplied.

  Res.Df Df     F Pr(>F)
1    534
2    533  1 3e-04  0.986
```

8. We again follow the steps from lecture 8 (slides 23 and 24 summarises) for estimating a nonlinear marginal effect, its standard error, and 95% CI.

   - Starting with income_scale=3 (or income=$30,000), the elasticity at this income level can be computed from the Reg (1) results as dY=0.385 + 0.046 x 3 = 0.523.

   - Computing the standard error now, let b1 be the coefficient on log_black and b2 be the coefficient on log_black_income. Holding other factors fixed, the predicted elasticity at income_scale=2 from a one unit change in black is then b1-b2 x 3. Testing the joint null hypothesis that b1 - b2 x 3 = 0 yields an F-statistic of 184.09 with 1 and 534 df. Therefore, the standard error of the elasticity at income_scale=3 is abs(0.523)/sqrt(184.09)=0.039.

   - Finally, the 95% CI for the elasticity at income_scale=2 is computed in the usual way as 95% CI = [0.523-1.96 x 0.039, 0.523 + 1.96 x 0.039]=[0.447,0.599].

We can compute the elasticity, its standard error, and 95% CI for income_scale=5 in the exact same way (see the as3.R code for details) and obtain an elasticity of 0.615, standard error of 0.023, and 95% CI of [0.570,0.660].

In words, we find that in states with higher income levels, the sensitivity (or elasticity) of the robbery rate with respect to the share of the black population is higher. Along this crime-based dimension, this suggests that higher-income states tend to exhibit higher degrees of discrepancy between black and white citizens.

9. As per the question on the assignment, full marks for the R code will be given it is as clear as the code in as2.R (or better!).

10. There is no “right” answer, submissions are assessed based on their clarity and thoughtfulness in describing research design, data, and econometric model.