Advanced Econometrics - Assignment 3

Problem 1:

1a)

First column:

Model: Y_it = alpha + Beta(X_it) + u_it
Formula: vio_it = alpha + beta(shall_it) + u_it
R Code: regpols <- lm(vio~shall,data=mydata)

Second Column:

Model: Y_it = beta(X_it) + alpha_i + u_it
Formula: vio_it = beta(shall_it) + stateid + u_it
R Code: regfes <- lm (vio~shall+factor(stateid),data=mydata)

Third Column:

Model: Y_it = beta(X_it) + alpha_i + lambda_t + u_it
Formula: vio_it = beta(shall_it) + stateid + lambda(year) + u_it
R Code: regfesy <- lm(vio~shall+factor(year)+factor(stateid),data=mydata)

Fourth Column:

Model: Y_it = beta(X_it) + alpha_i + lambda_t + gamma_i(t) + u_it
Formula: vio_it = beta(shall_it) + year + stateid + gamma(dummy) + u_it
R Code: regfesyt <- lm(vio~shall+factor(year)+factor(stateid)+factor(stateid):year,data=mydata)

Fifth Column:

Model: beta(X_it) + gamma_1((W_it)^(1)) + gamma_2((W_it)^(2)) + … + alpha_i + lambda_t + u_it
Formula: vio_it = beta(shall_it) + gamma_1(density) + gamma_2(avginc) + gamma_3(incarc_rate) + gamma_4(pm1029) + stateid + year + u_it
R Code: regfesytmulti <- lm(vio~shall+density+avginc+incarc_rate+pm1029+factor(year)+factor(stateid),data=mydata)

1b)

First Column:

Model: Y_it = alpha + beta(X_it) + u_it
Formula: vio_it = alpha + beta(shall_it) + u_it
R Code: pregpols <- plm(vio~shall,model="pooling",data=mydatap)

Second Column:

Model: Y_it = beta(X_it – Xbar_i) + u_it – ubar_i
Formula: vio_it = beta(shall_it – shall-bar_i) + u_it – ubar_i
R Code: pregfes <- plm(vio~shall,model="within",data=mydatap)

Third Column:

Model: Y_it = beta(X_it) + alpha_i + lambda_t + u_it
Formula: vio_it = beta(shall_it) + stateid + year + u_it

R Code: pregfesy <- plm(vio~shall,effect="twoway",model="within",data=mydatap)

Fourth Column:

Formula: vio_it = beta(shall_it) + gamma_1(density) + gamma_2(avginc) + gamma_3(incarc_rate) + gamma_4(pm1029) + stateid + year + u_it

R Code: pregfesymulti <- plm(vio~shall+density+avginc+incarc_rate+pm1029,effect="twoway",model="within",data=mydatap)

1c) The standard errors are different between the two tables because on table 1, the standard errors are "computed on the assumptions of homoskedasticity and no serial correlation!" (Week 6, Slide 24). It is more than likely that there is heteroskedastic overlap between several variables.

1d) According to table 1a, there is a mostly negative effect of "shall" carry laws on violent crime rates based on regression 1. However, when we control for fixed effects, the effect becomes positive as seen in regression 2.

If we include time-varying variables, then the effect is no longer significant like in regression 1 and 2 of table 1. From this information, we gather that if the regression does not include time-varying controls, we cannot assume there are significant causal effects in either the fixed effects model or the pooled OLS model.

Problem 2:

2a) The model that minimizes AIC, BIC, and CV is regression of Sales on TV and Radio. The AIC is 212.819, BIC is 226.012, CV is 2.911. It also has the highest R^2 value of 0.8962.

2b) The LASSO results are:

```
> X = model.matrix(Sales~.,mydata)
> y = mydata$Sales
> lassoreg = cv.glmnet(X, y, alpha = 1)
> predict(lassoreg, type = "coefficients")
6 x 1 sparse Matrix of class "dgCMatrix"
                      1
(Intercept) 4.28174584
(Intercept) .
X           .
TV          0.04092247
Radio       0.16005036
Newspaper   .
```

Based on these results, the variables used are TV and Radio. These are only indicators of which variables to use, not unbiased regression coefficients.

2c) Using the LASSO method, the output tells us to use the regression with the variable's TV and Radio. However, these coefficients are bias and not indicative of the actual coefficients when running the regression of Sales on TV and Radio.

Based on our comparison of all the regressions and the possible combinations, the regression of Sales on TV and Radio yields us:

```
> CV(regt)
         CV         AIC        AICc         BIC       AdjR2
 10.7410876 476.5159143 476.6383632 486.4108664   0.6099148
> CV(regr)
         CV         AIC        AICc         BIC       AdjR2
 18.4856639 585.0983672 585.2208162 594.9933193   0.3286589
> CV(regn)
          CV          AIC         AICc          BIC        AdjR2
 26.27078080 655.09593833 655.21838730 664.99089042   0.04733317
> CV(regtr)
         CV         AIC        AICc         BIC       AdjR2
  2.9106758 212.8186854 213.0238136 226.0119549   0.8961505
> CV(regtn)
         CV         AIC        AICc         BIC       AdjR2
  9.8975987 460.2027523 460.4078805 473.3960218   0.6422399
> CV(regrn)
         CV         AIC        AICc         BIC       AdjR2
 18.7544100 586.8968402 587.1019684 600.0901097   0.3259306
> CV(regall)
         CV         AIC        AICc         BIC       AdjR2
  2.9468998 214.7868226 215.0961010 231.2784094   0.8956373
```

Based on the information in the LASSO analysis and the CV analysis, in order to create the most accurate regression, only the variables TV and Radio are used!

Problem 3:

a)

```
> stargazer(regsp, regsd, regsr, type = "text")

===============================================================================================
                                        Dependent variable:
                    ---------------------------------------------------------------------------
                                SP500                      SP500D                  SP500R
                                 (1)                        (2)                     (3)
-----------------------------------------------------------------------------------------------
SP500L                         0.999***
                               (0.001)

SP500D1                                                  -0.078***
                                                         (0.020)

SP500RL                                                                          -0.00000
                                                                                 (0.00000)

Constant                        1.285                      0.281                   0.001
                               (1.381)                    (0.322)                 (0.001)

-----------------------------------------------------------------------------------------------
Observations                   2,517                      2,516                   2,517
R2                             0.998                      0.006                   0.0005
Adjusted R2                    0.998                      0.006                   0.0001
Residual Std. Error       16.208 (df = 2515)         16.163 (df = 2514)      0.013 (df = 2515)
F Statistic        1,160,097.000*** (df = 1; 2515) 15.460*** (df = 1; 2514) 1.213 (df = 1; 2515)
===============================================================================================
Note:                                                          *p<0.1; **p<0.05; ***p<0.01
>
```

b) The first regression, SP500 on its lag SP550L, is considered spurious. This is because the data is trending together and needs to be corrected in order to be considered. A regression cannot be performed unless the variable is considered stationary.

c) Based on the regression of SP500R on its lag SP500RL, there is a slight negative auto-correlation. The returns coefficient is -8.2923e-07. The t-statistic is also less than 1.96, which indicates that is it not a significant relationship.