GISD Seminar Markdown

Simon Kolbeck

12/9/2019

Here I am calling in the data. Please see the manuscript for a description of where the data was obtained.

```
setwd("~/Clearance Project/R")
#install.packages("readxl")
#library(readxl)
cd <- read.csv("ClearanceFD.csv")
names(cd)
```

In the next chunk of code, I am using linear interpolation to estimate the missing values of each city's annual population, black population (as a percentage) and the total number of sworn full-time police offers. Using the match function, I pull the interpolated values back into the master data "cd", overriding the missing values.

```
##INTERPOLATIONS ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
# Annual Pop 1990-2000 Interpolating++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
IP <- cd %>%
  group_by(City) %>%
  mutate(valueIpol = approx(Year, annual_pop, Year,
                            method="linear",
                            rule=1, f=0,
                            ties = mean)$y)
#ipop = interpolated population
#Pulling interpolated population into main data:
cd$annual_pop <- IP$valueIpol[match(cd$cy,IP$cy)]
#write.csv(IP,file="pop_interpolation.csv",row.names=FALSE)

#Interpolating percent black:
IB <- cd %>%
  group_by(City) %>%
  mutate(valueIpol = approx(Year, bl, Year,
                            method="linear",
                            rule=2, f=0,
                            ties = mean)$y)
cd$bl <- IB$valueIpol[match(cd$cy,IB$cy)]
#write.csv(IB,file="black_interpolation.csv",row.names=FALSE)

#Interpolating sworn officers
ISO<-cd %>%
  group_by(City) %>%
  mutate(valueIpol = approx(Year, so, Year,
                            method = "linear",
                            rule = 2, f = 0,
                            ties = mean)$y)
cd$so <- ISO$valueIpol[match(cd$cy,ISO$cy)]
#write.csv(ISO, file="pd.csv",row.names = FALSE)
```

```
cd$clr_lag <- as.numeric(cd$clr_lag)
cd$clr_r <- as.numeric(cd$clr_r)
cd$hom <- as.numeric(cd$hom)
lapply(cd,class)
```

The next chunk of code calculates the annual homicide rate per 100,000 pop. and the annual number of sworn police officers per 100,000 pop. In addition, using the "subset" function, I remove all variables from the year 1990 (first year in the series) because the primary dependent variables is the lagged clearance rate.

```
#Calculating annual rates, etc. using interpolated population
#Annual Homicide per 100,000 population:
for(i in 1:nrow(cd)){
  cd$hom_rate[i] <- ((cd$hom[i]/cd$annual_pop[i])*100000)
}
#Annual sworn officers per 100,000 population:
cd$so_pop <- NA
for(i in 1:nrow(cd)){
  cd$so_pop[i] <- ((cd$so[i]/cd$annual_pop[i])*100000)
}

#Getting rid of Year 1990:
cd <- cd[which(cd$Year!=1990), ]
```

For some reason, R turned all missing values of the lagged clearance measure into "3". The code below corrects this.

```
for(i in 1:nrow(cd)){
  if(cd$clr_lag[i]==3){
    cd$clr_lag[i] <- NA
  }
}
```

Next, I run my models with the annual homicide rate as the dependent variable. The first model contains a factor for each city coded as +factor(City). Thereby, I control for unique city effects. However, I also use the "plm" package to run a fixed effects model with the same variables. The results are effectively identical, apart from the R-squared, which is much higher in the factor model. I also check for robust standard errors and create a LaTex regression table using the "stargazer" function.

```
#Model with lagged clearance rate
cm1.1 <- lm(hom_rate ~ clr_lag + bl + unemp + so + factor(City), data=cd)
summary(cm1.1)
#Model with non-lagged clearance rate
#summary(lm(hom_rate ~ clr_r + ibl + unemp + factor(City), data=cd))

#hom rate as dependant variable
summary(lm(clr ~ hom_rate + bl + unemp + so_pop + factor(City), data=cd2))

install.packages("sandwich")
library(sandwich)
install.packages("lmtest")
library(lmtest)
```

```
vcv <- vcovHC(cm1.1, type = "HC1" )
coeftest(cm1.1, vcv)

#Model using plm function
install.packages("plm")
library(plm)
cm2.1 <- plm(hom_rate ~ clr_lag + bl + unemp + so_pop , data=cd, index=c("City","Year"), model="within"
summary(cm2.1)
#With homicides as independent variable:
cm2.2 <- plm(clr_r ~ hom_rate + bl + unemp + so_pop, data=cd, index=c("City","Year"), model="within")
summary(cm2.2)
#STARGAZER Conversion to LaTex:
install.packages('stargazer')
library(stargazer)
stargazer(cm2.1, title="Results",align=TRUE)
```

The code below creates a scatterplot of the bivariate relationship between annual homicide rate and the clearance rate.

```
install.packages("ggplot2")
library(ggplot2)
cp1 <- ggplot(data=cd, mapping = aes(x=clr_lag, y=hom_rate))
cp1 +  geom_point(color = " blue") +
  geom_smooth(method="lm", color="orange") +
  #geom_text(data = subset(cnrm,mean_loghom > 3), mapping = aes(label = X)) +
  theme_minimal() +
  xlab("Lagged Clearance Rate") +
  ylab("Homicide Rate") +
  ggtitle("Bivariate Relationship Between Homicide Clearance Rates and Homicide Rates")
```