BT2101 GA6 Group 67 Submission

Lo Zhi Hao

2022-11-06

1 Instrumental Variable Regression using R

Please use the movies data set to answer the following questions. Please carefully read the document of data description.

Does viewing a violent movie lead to violent behavior? If so, the incidence of violent crimes, such as assaults, should rise following the release of a violent movie. Alternatively, movie viewing may substitute for other activities (such as alcohol consumption) that lead to violent behavior, so that assaults should fall when more people are attracted to the cinema (rather than alcohol consumption). The data set movies contains data on the number of assaults and movie attendance for 516 weekends from 1995 through 2004. It includes weekend attendance for violent movies (such as Hannibal), mildly violent movies (such as SpiderMan), and nonviolent movies (such as Finding Nemo). Additionally, the data set includes a count of assaults for the same weekend in a subset of counties. Lastly, the data set includes indicators for the year, month, whether the weekend is a holiday, and various weather measures.

```
## Setting up the environment for further studies

## install.packages("wooldridge")
## install.packages("dplyr")
## install.packages("ggplot2")
## install.packages("ivreg", dependencies = TRUE)
## install.packages("tidyverse")

library(wooldridge)
library(dplyr)
library(knitr)
library(corrplot)
library(ggplot2)
library(ivreg)
library(readxl)

## loading the dataset
movies <- read_excel("GA6-movies.xlsx")
summary(movies)
head(movies)
```

a. Regress ln assaults onto attend v, using the year and month indicators, the weather and holiday indicators as control variables (denoted as the basic model). Based on this regression, is viewing a strongly violent movie correlate with the number of assaults? Interpret the coefficient estimates (Notes: When interpreting estimates, always include the precision of these estimates in your discussion).

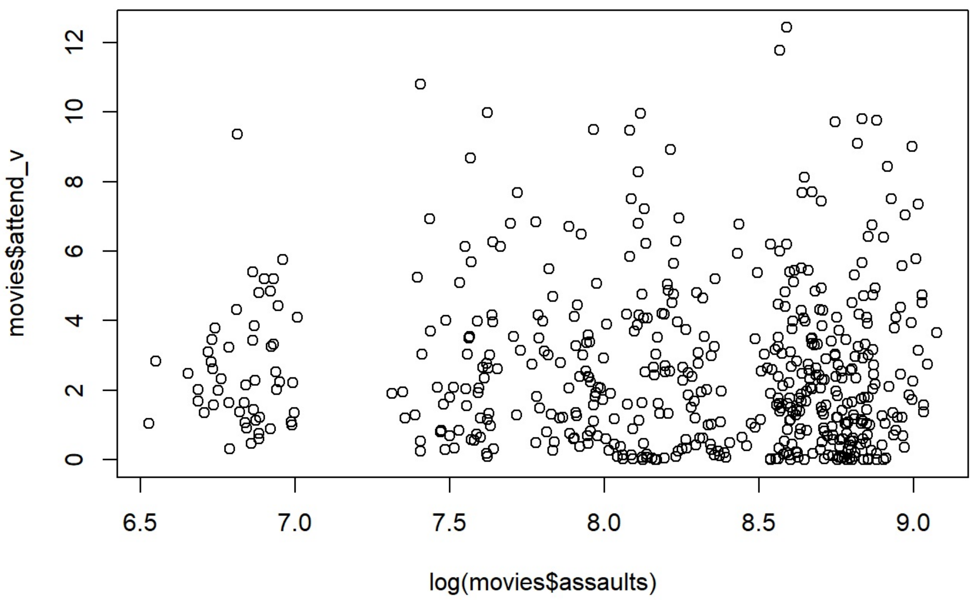
```
## mutating variables

movies <- movies %>%
  mutate(yr = ifelse(year1 == 1, 1,
                     ifelse(year2 == 1, 2,
                            ifelse(year3 == 1, 3,
                                   ifelse(year4 == 1, 4,
                                          ifelse(year5 == 1, 5,
                                                 ifelse(year6 == 1, 6,
                                                        ifelse(year7 == 1, 7,
                                                               ifelse(year8 == 1, 8,
                                                                      ifelse(year9 == 1, 9, 10)))))))))) %>%
  mutate(month = ifelse(month1 == 1, 1,
                     ifelse(month2 == 1, 2,
                            ifelse(month3 == 1, 3,
                                   ifelse(month4 == 1, 4,
                                          ifelse(month5 == 1, 5,
                                                 ifelse(month6 == 1, 6,
                                                        ifelse(month7 == 1, 7,
                                                               ifelse(month8 == 1, 8,
                                                                      ifelse(month9 == 1, 9,
                                                                             ifelse(month10 == 1, 10,
                                                                                    ifelse(month11 == 1, 11, 12))
))))))))))

## creating the linear model

basic.model <- lm(log(assaults) ~ attend_v + as.factor(yr) + as.factor(month) + h_chris + h_newyr + h_easter + h_
july4 + h_mem + h_labor + w_rain + w_snow + w_maxa + w_maxb + w_maxc + w_mina + w_minb + w_minc, data = movies)
summary(basic.model)
```

```
##
## Call:
## lm(formula = log(assaults) ~ attend_v + as.factor(yr) + as.factor(month) +
##     h_chris + h_newyr + h_easter + h_july4 + h_mem + h_labor +
##     w_rain + w_snow + w_maxa + w_maxb + w_maxc + w_mina + w_minb +
##     w_minc, data = movies)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.31675 -0.02527  0.00034  0.02463  0.19730
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)         6.8787365  0.0151775 453.219  < 2e-16 ***
## attend_v           -0.0008530  0.0008919  -0.956  0.33936
## as.factor(yr)2      0.6958578  0.0086678  80.281  < 2e-16 ***
## as.factor(yr)3      1.0083695  0.0086703 116.301  < 2e-16 ***
## as.factor(yr)4      1.2162494  0.0087517 138.973  < 2e-16 ***
## as.factor(yr)5      1.3816193  0.0087480 157.936  < 2e-16 ***
## as.factor(yr)6      1.6805043  0.0085976 195.463  < 2e-16 ***
## as.factor(yr)7      1.8277952  0.0086706 210.804  < 2e-16 ***
## as.factor(yr)8      1.8816248  0.0086934 216.442  < 2e-16 ***
## as.factor(yr)9      1.9370691  0.0085639 226.190  < 2e-16 ***
## as.factor(yr)10     2.0611838  0.0089017 231.551  < 2e-16 ***
## as.factor(month)2  -0.0050253  0.0097632  -0.515  0.60699
## as.factor(month)3   0.0180803  0.0102375   1.766  0.07802 .
## as.factor(month)4   0.0108111  0.0125102   0.864  0.38792
## as.factor(month)5   0.0067936  0.0146356   0.464  0.64273
## as.factor(month)6  -0.0382115  0.0158567  -2.410  0.01634 *
## as.factor(month)7  -0.0490938  0.0178493  -2.750  0.00618 **
## as.factor(month)8  -0.0415764  0.0171478  -2.425  0.01569 *
## as.factor(month)9  -0.0004429  0.0151478  -0.029  0.97668
## as.factor(month)10  0.0052716  0.0126945   0.415  0.67813
## as.factor(month)11 -0.0511186  0.0107731  -4.745 2.75e-06 ***
## as.factor(month)12 -0.0286075  0.0100976  -2.833  0.00480 **
## h_chris            -0.1020299  0.0237133  -4.303 2.05e-05 ***
## h_newyr             0.2240654  0.0228015   9.827  < 2e-16 ***
## h_easter           -0.0414692  0.0148326  -2.796  0.00538 **
## h_july4             0.0353795  0.0206916   1.710  0.08794 .
## h_mem              -0.0144231  0.0145817  -0.989  0.32310
## h_labor             0.0261436  0.0145328   1.799  0.07266 .
## w_rain             -0.0339778  0.0130830  -2.597  0.00969 **
## w_snow             -0.0674832  0.0302649  -2.230  0.02623 *
## w_maxa              0.1091011  0.0137728   7.921 1.65e-14 ***
## w_maxb              0.1058065  0.0188005   5.628 3.11e-08 ***
## w_maxc              0.0273434  0.0709579   0.385  0.70015
## w_mina             -0.3239640  0.0401277  -8.073 5.55e-15 ***
## w_minb             -0.1691211  0.0275710  -6.134 1.79e-09 ***
## w_minc             -0.1269039  0.0170766  -7.431 4.96e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04282 on 480 degrees of freedom
## Multiple R-squared:  0.9957, Adjusted R-squared:  0.9954
## F-statistic:  3177 on 35 and 480 DF,  p-value: < 2.2e-16
```

```
## plotting correlation to
cor(log(movies$assaults), movies$attend_v)
```

```
## [1] -0.02480429
```

```
plot(log(movies$assaults), movies$attend_v)
```



The relationship is as follows:

`log(assaults)` = 6.8787365 - 0.0008530 x `attend_v` + other independent control variables, including year, month, holiday and weather

Multiple R squared is 0.9957, and Adjusted R squared is 0.9954.

When there are no viewing of strongly violent movies, for year 1995 month January without any considering any fraction of weather and with no holiday, the number of assaults and intimidation in a subset of U.S. counties is e^6.8787365. For coefficient of viewing of strongly violent movies, it is -0.0008530, which suggests that a one million increase in viewing of strongly violent movie in year 1995 January without any considering any fraction of weather and without holiday is associated with a 0.0853% decrease in the number of assaults and intimidation in a subset of U.S. counties. This coefficient is statistically insignificant (p-value = 0.33936 > 0.05), which suggests that we have insufficient statistical evidence to reject the null hypothesis and conclude that the relationship is significantly different from 0. Essentially, we cannot conclude that the viewing of strongly violent movie has any association with the number of assaults and intimidation in a subset of U.S. counties.

b. It is difficult to control for all confounders that bias the relationship between assaults and movie attendance. However, the data set includes a potential instrument variable (IV): pr attend v. This is a predicted value for the film's attendance in a given weekend based only on historical attendance patterns (i.e., it does not utilize information from that same weekend on which the prediction is made). For example, if a film's attendance is high in the first week of its release, then this can be used to predict that its attendance is also high in the second week of its release. Please explain whether pr attend v is a valid instrument variable according to your understanding.

```
## finding correlation between pr_attend_v and attend_v
cor(movies$pr_attend_v, movies$attend_v)
```

```
## [1] 0.9642839
```

For a variable to be a instrumental variable, it has to satisfy both instrument relevance and instrument exogeneity.

For pr_attend_v, as seen by the correlation computed above, the correlation between pr_attend_v and attend_v is very high at 0.964. From a logical standpoint, this is reasonable as pr_attend_v is computed by taking into account the predicted value for the violent rated movie based on historical attendance pattern. Therefore, as pr_attend_v and attend_v are highly correlated and the correlation between these two variables are reasonable, it is safe to say that pr_attend_v satisfies instrument relevance.

For pr_attend_v, as shown by the definition of pr_attend_v, it is a value that is predicted based on the historical attendance data for every week, and thus it is only correlated to assault numbers through the the actual attendance of violent movies. Therefore, it is not affected by the fraction of weather for the week and whether the weekend is a holiday or not, as it utilises historical data that precedes both of these. At the same time, as it utilises historical data, it is therefore not correlated with any of the years and months indicator, and we can suggest with confidence that the pr_attend_v is not affected in a significant way by years and months as well. Therefore, as pr_attend_v is not likely to be correlated to every other control variable in the model, we can safely conclude that it satisfies the instrument exogeneity requirement as well.

Therefore, as it satisfies both instrument relevance and instrument exogeneity, we can confidently conclude that pr_attend_v is a valid instrument variable according to our understanding.

c. Estimate a model that uses pr attend v as an instrument variable to replace attend v in the model in Question (a) Using the lm() command in R to execute 2SLS manually. Interprete the coeficient estimates. Compared to the model in Question (a), explain the change in coefficient estimates, if any.

```r
## variables:
## dependent variable : log(assaults)
## endogeneous variable : attend_v
## exogeneous variable : pr_attend_v

## finding predicted values:
tempLm <- lm(attend_v ~ pr_attend_v + as.factor(yr) + as.factor(month) + h_chris + h_newyr + h_easter + h_july4 +
h_mem + h_labor + w_rain + w_snow + w_maxa + w_maxb + w_maxc + w_mina + w_minb + w_minc, data = movies)
summary(tempLm)
```

```
##
## Call:
## lm(formula = attend_v ~ pr_attend_v + as.factor(yr) + as.factor(month) +
##     h_chris + h_newyr + h_easter + h_july4 + h_mem + h_labor +
##     w_rain + w_snow + w_maxa + w_maxb + w_maxc + w_mina + w_minb +
##     w_minc, data = movies)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.3347 -0.2692 -0.0357  0.1963  3.8132
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -0.195486   0.209515  -0.933  0.35127
## pr_attend_v         0.942324   0.012027  78.352  < 2e-16 ***
## as.factor(yr)2      0.235996   0.119516   1.975  0.04889 *
## as.factor(yr)3      0.316209   0.119305   2.650  0.00830 **
## as.factor(yr)4      0.313239   0.120793   2.593  0.00980 **
## as.factor(yr)5      0.191662   0.120526   1.590  0.11245
## as.factor(yr)6      0.160550   0.118445   1.355  0.17590
## as.factor(yr)7      0.355382   0.119754   2.968  0.00315 **
## as.factor(yr)8      0.262724   0.120365   2.183  0.02954 *
## as.factor(yr)9      0.313059   0.117958   2.654  0.00822 **
## as.factor(yr)10     0.324759   0.122595   2.649  0.00834 **
## as.factor(month)2   0.254995   0.134376   1.898  0.05835 .
## as.factor(month)3   0.111029   0.141029   0.787  0.43151
## as.factor(month)4   0.009237   0.172408   0.054  0.95729
## as.factor(month)5   0.002092   0.201705   0.010  0.99173
## as.factor(month)6  -0.186692   0.218592  -0.854  0.39349
## as.factor(month)7   0.171976   0.245864   0.699  0.48459
## as.factor(month)8   0.353485   0.236087   1.497  0.13498
## as.factor(month)9   0.317643   0.208689   1.522  0.12865
## as.factor(month)10  0.082225   0.174908   0.470  0.63849
## as.factor(month)11  0.151309   0.148353   1.020  0.30828
## as.factor(month)12  0.116264   0.139201   0.835  0.40401
## h_chris            -0.160787   0.326842  -0.492  0.62299
## h_newyr             0.336696   0.314257   1.071  0.28453
## h_easter            0.186345   0.204567   0.911  0.36279
## h_july4            -0.315365   0.285060  -1.106  0.26915
## h_mem              -0.064963   0.200974  -0.323  0.74666
## h_labor            -0.012914   0.200282  -0.064  0.94861
## w_rain             -0.146378   0.180389  -0.811  0.41751
## w_snow              0.908390   0.416803   2.179  0.02979 *
## w_maxa             -0.041727   0.189820  -0.220  0.82610
## w_maxb             -0.036809   0.259099  -0.142  0.88709
## w_maxc             -0.262417   0.977833  -0.268  0.78853
## w_mina             -0.832017   0.552898  -1.505  0.13303
## w_minb             -0.442310   0.379875  -1.164  0.24486
## w_minc              0.202497   0.235264   0.861  0.38982
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5901 on 480 degrees of freedom
## Multiple R-squared:  0.9375, Adjusted R-squared:  0.9329
## F-statistic: 205.7 on 35 and 480 DF,  p-value: < 2.2e-16
```

```r
attend_hat <- fitted.values(tempLm)
movies <- movies %>% mutate(attend_hat = attend_hat)

## running the linear model
basic.model2 <- lm(log(assaults) ~ attend_hat + as.factor(yr) + as.factor(month) + h_chris + h_newyr + h_easter +
h_july4 + h_mem + h_labor + w_rain + w_snow + w_maxa + w_maxb + w_maxc + w_mina + w_minb + w_minc, data = movies)
summary(basic.model2)
```

```
##
## Call:
## lm(formula = log(assaults) ~ attend_hat + as.factor(yr) + as.factor(month) +
##     h_chris + h_newyr + h_easter + h_july4 + h_mem + h_labor +
##     w_rain + w_snow + w_maxa + w_maxb + w_maxc + w_mina + w_minb +
##     w_minc, data = movies)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.31661 -0.02496  0.00093  0.02441  0.19776
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)         6.8791439  0.0151755 453.307  < 2e-16 ***
## attend_hat         -0.0010973  0.0009256  -1.185  0.23641
## as.factor(yr)2      0.6958397  0.0086634  80.320  < 2e-16 ***
## as.factor(yr)3      1.0085237  0.0086673 116.359  < 2e-16 ***
## as.factor(yr)4      1.2161845  0.0087475 139.032  < 2e-16 ***
## as.factor(yr)5      1.3816766  0.0087437 158.020  < 2e-16 ***
## as.factor(yr)6      1.6805752  0.0085935 195.564  < 2e-16 ***
## as.factor(yr)7      1.8277099  0.0086666 210.891  < 2e-16 ***
## as.factor(yr)8      1.8813257  0.0086944 216.385  < 2e-16 ***
## as.factor(yr)9      1.9371468  0.0085599 226.305  < 2e-16 ***
## as.factor(yr)10     2.0612741  0.0088976 231.666  < 2e-16 ***
## as.factor(month)2  -0.0048340  0.0097602  -0.495  0.62063
## as.factor(month)3   0.0182388  0.0102335   1.782  0.07534 .
## as.factor(month)4   0.0108440  0.0125039   0.867  0.38624
## as.factor(month)5   0.0066987  0.0146285   0.458  0.64722
## as.factor(month)6  -0.0381273  0.0158488  -2.406  0.01652 *
## as.factor(month)7  -0.0485345  0.0178493  -2.719  0.00678 **
## as.factor(month)8  -0.0412551  0.0171422  -2.407  0.01648 *
## as.factor(month)9  -0.0003299  0.0151405  -0.022  0.98263
## as.factor(month)10  0.0055318  0.0126908   0.436  0.66311
## as.factor(month)11 -0.0508688  0.0107706  -4.723 3.06e-06 ***
## as.factor(month)12 -0.0286646  0.0100926  -2.840  0.00470 **
## h_chris            -0.1019405  0.0237014  -4.301 2.06e-05 ***
## h_newyr             0.2240793  0.0227899   9.832  < 2e-16 ***
## h_easter           -0.0416665  0.0148264  -2.810  0.00515 **
## h_july4             0.0351880  0.0206820   1.701  0.08952 .
## h_mem              -0.0143619  0.0145744  -0.985  0.32491
## h_labor             0.0261129  0.0145254   1.798  0.07285 .
## w_rain             -0.0338525  0.0130770  -2.589  0.00993 **
## w_snow             -0.0671544  0.0302514  -2.220  0.02689 *
## w_maxa              0.1091570  0.0137659   7.929 1.56e-14 ***
## w_maxb              0.1058051  0.0187910   5.631 3.06e-08 ***
## w_maxc              0.0266674  0.0709251   0.376  0.70709
## w_mina             -0.3241934  0.0401079  -8.083 5.18e-15 ***
## w_minb             -0.1693078  0.0275576  -6.144 1.69e-09 ***
## w_minc             -0.1267221  0.0170689  -7.424 5.21e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04279 on 480 degrees of freedom
## Multiple R-squared:  0.9957, Adjusted R-squared:  0.9954
## F-statistic:  3180 on 35 and 480 DF,  p-value: < 2.2e-16
```

The relationship is as follows:

`log(assaults)` = 6.8791439 - 0.0010973 x `attend_hat` + other independent control variables, including year, month, holiday and weather

Multiple R squared is 0.9957, and Adjusted R squared is 0.9954.

When there are no viewing of strongly violent movies, for year 1995 month January without any considering any fraction of weather and with no holiday, the number of assaults and intimidation in a subset of U.S. counties is e^6.8791439. For coefficient of viewing of strongly violent movies, it is -0.0010973, which suggests that a one million increase in viewing of strongly violent movie in year 1995 January without any considering any fraction of weather and without holiday is associated with a 0.109% decrease in the number of assaults and intimidation in a subset of U.S. counties. This coefficient is statistically insignificant (p-value = 0.23641 > 0.05), which suggests that we have insufficient statistical evidence to reject the null hypothesis and conclude that the relationship is significantly different from 0. Essentially, we cannot conclude that the viewing of strongly violent movie has any association with the number of assaults and intimidation in a subset of U.S. counties.

The intercept goes from 6.8787365 to 6.8791439, marked a 0.0004074 difference between models in (a) and (c). And they are both statistical significance with p-values of 2e-16. Compared to model in part (a), despite everything else remain constant other than substituting in attend_hat for attend_v, we notice that there is a slight change in coefficient from -0.0008530 to -0.0010973 for attend_v and attend_hat respectively. The change in both intercept and coefficient for attend_v might be due to the part of viewings of strongly violent movies that are deviated and biased by non-measured confounding variables being removed from the linear model through the use of instrumental variable pr_attend_v. The p-value still remained statistically insignificant despite the change in coefficient and variables, which suggests that we still are unable to conclude that the viewing of strongly violent movie has any association with the number of assaults and intimidation in a subset of U.S. counties.

d. Using the ivreg() function in R, estimate a similar IV regression model as shown in Question (c). Compare this model to the model in Question (c) AND the basic model in Question (a), explain the change in coefficient estimates, if any

```
## running ivreg()
basic.model3 <- ivreg(log(assaults) ~ attend_v + as.factor(yr) + as.factor(month) + h_chris + h_newyr + h_easter
+ h_july4 + h_mem + h_labor + w_rain + w_snow + w_maxa + w_maxb + w_maxc + w_mina + w_minb + w_minc | as.factor(y
r) + as.factor(month) + h_chris + h_newyr + h_easter + h_july4 + h_mem + h_labor + w_rain + w_snow + w_maxa + w_m
axb + w_maxc + w_mina + w_minb + w_minc + pr_attend_v , data = movies)
summary(basic.model3)
```

```
##
## Call:
## ivreg(formula = log(assaults) ~ attend_v + as.factor(yr) + as.factor(month) +
##     h_chris + h_newyr + h_easter + h_july4 + h_mem + h_labor +
##     w_rain + w_snow + w_maxa + w_maxb + w_maxc + w_mina + w_minb +
##     w_minc | as.factor(yr) + as.factor(month) + h_chris + h_newyr +
##     h_easter + h_july4 + h_mem + h_labor + w_rain + w_snow +
##     w_maxa + w_maxb + w_maxc + w_mina + w_minb + w_minc + pr_attend_v,
##     data = movies)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.3169309 -0.0250011  0.0005169  0.0245514  0.1970052
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          6.8791439  0.0151844 453.040  < 2e-16 ***
## attend_v            -0.0010973  0.0009262  -1.185  0.23668
## as.factor(yr)2       0.6958397  0.0086685  80.273  < 2e-16 ***
## as.factor(yr)3       1.0085237  0.0086724 116.291  < 2e-16 ***
## as.factor(yr)4       1.2161845  0.0087526 138.951  < 2e-16 ***
## as.factor(yr)5       1.3816766  0.0087488 157.927  < 2e-16 ***
## as.factor(yr)6       1.6805752  0.0085985 195.449  < 2e-16 ***
## as.factor(yr)7       1.8277099  0.0086717 210.767  < 2e-16 ***
## as.factor(yr)8       1.8813257  0.0086995 216.257  < 2e-16 ***
## as.factor(yr)9       1.9371468  0.0085649 226.172  < 2e-16 ***
## as.factor(yr)10      2.0612741  0.0089028 231.530  < 2e-16 ***
## as.factor(month)2   -0.0048340  0.0097659  -0.495  0.62083
## as.factor(month)3    0.0182388  0.0102395   1.781  0.07551 .
## as.factor(month)4    0.0108440  0.0125112   0.867  0.38652
## as.factor(month)5    0.0066987  0.0146371   0.458  0.64741
## as.factor(month)6   -0.0381273  0.0158581  -2.404  0.01658 *
## as.factor(month)7   -0.0485345  0.0178598  -2.718  0.00682 **
## as.factor(month)8   -0.0412551  0.0171523  -2.405  0.01654 *
## as.factor(month)9   -0.0003299  0.0151494  -0.022  0.98264
## as.factor(month)10   0.0055318  0.0126982   0.436  0.66330
## as.factor(month)11  -0.0508688  0.0107769  -4.720 3.10e-06 ***
## as.factor(month)12  -0.0286646  0.0100985  -2.838  0.00472 **
## h_chris             -0.1019405  0.0237154  -4.298 2.08e-05 ***
## h_newyr              0.2240793  0.0228033   9.827  < 2e-16 ***
## h_easter            -0.0416665  0.0148351  -2.809  0.00518 **
## h_july4              0.0351880  0.0206942   1.700  0.08971 .
## h_mem               -0.0143619  0.0145830  -0.985  0.32520
## h_labor              0.0261129  0.0145339   1.797  0.07301 .
## w_rain              -0.0338525  0.0130847  -2.587  0.00997 **
## w_snow              -0.0671544  0.0302692  -2.219  0.02698 *
## w_maxa               0.1091570  0.0137740   7.925 1.61e-14 ***
## w_maxb               0.1058051  0.0188020   5.627 3.12e-08 ***
## w_maxc               0.0266674  0.0709668   0.376  0.70725
## w_mina              -0.3241934  0.0401315  -8.078 5.36e-15 ***
## w_minb              -0.1693078  0.0275738  -6.140 1.73e-09 ***
## w_minc              -0.1267221  0.0170790  -7.420 5.37e-13 ***
##
## Diagnostic tests:
##                   df1 df2 statistic p-value
## Weak instruments    1 480   6138.98  <2e-16 ***
## Wu-Hausman          1 479      0.96   0.328
## Sargan              0  NA        NA      NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04282 on 480 degrees of freedom
## Multiple R-Squared: 0.9957,  Adjusted R-squared: 0.9954
## Wald test:  3176 on 35 and 480 DF,  p-value: < 2.2e-16
```

The relationship is as follows:

`log(assaults)` = 6.8791439 - 0.0010973 x `attend_v` + other independent control variables, including year, month, holiday and weather

Multiple R squared is 0.9957, and Adjusted R squared is 0.9954.

When there are no viewing of strongly violent movies, for year 1995 month January without any considering any fraction of weather and with no holiday, the number of assaults and intimidation in a subset of U.S. counties is e^6.8791439. For coefficient of viewing of strongly violent movies, it is -0.0010973, which suggests that a one million increase in viewing of strongly violent movie in year 1995 January without any considering any fraction of weather and without holiday is associated with a 0.109% decrease in the number of assaults and intimidation in a subset of U.S. counties. This coefficient is statistically insignificant (p-value = 0.23641 > 0.05), which suggests that we have insufficient statistical evidence to reject the null hypothesis and conclude that the relationship is significantly different from 0. Essentially, we cannot conclude that the viewing of strongly violent movie has any association with the number of assaults and intimidation in a subset of U.S. counties.

The intercept goes from 6.8787365 to 6.8791439, marked a 0.0004074 difference between models in (a) and (d). And they are both statistical significance with p-values of 2e-16. Compared to model in part (a), we notice that there is a slight change in coefficient from -0.0008530 to -0.0010973 for attend_v in part (a) and attend_v in part (d) respectively. This might be due to the part of viewings of strongly violent movies that are deviated and biased by non-measured confounding variables being removed from the linear model through the use of the ivreg package and function. The p-value still remained statistically insignificant despite the change in coefficient and variables, which suggests that we still are unable to conclude that the viewing of strongly violent movie has any association with the number of assaults and intimidation in a subset of U.S. counties.

Compared to model in part (c), we can notice that the two models essentially produces the same results, with completely identical coefficients for attend_hat and attend_v in both models respectively. Further research actually suggests that these two models are essentially performing the same functions, and therefore produces highly similar results.