THE UNIVERSITY OF QUEENSLAND
AUSTRALIA
CREATE CHANGE

ECON2300 - Introductory Econometrics

Tutorial 10: Instrumental Variables Regression

Tutor: Francisco Tavares Garcia

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

R-Exercise 5 is available!

Dear ECON2300 Students,

R-Exercise 5 is now available in the "R-Exercises: Analysis of Data and Short Report" folder, which you can access via the Assessment tab.

The due date for R-Exercise 5 is **Friday, October 13, 2023, 4pm**

Please read all instructions carefully before commencing the R-Exercise. For convenience, a copy of the R-Exercise instructions has been presented below.

==

Instructions:

Please pay close attention to the number of decimal places required (if any) for each answer. The required number of decimal places may differ from question to question.

Avoid rounding during intermediate calculations where possible.

This R-Exercise is not timed. This means that you can open the R-Exercise and return to it as many times as you need to (provided that you do not click submit).

There is only one attempt for this R-Exercise.

The R-Exercise is marked out of 7, but will contribute 10% towards your final grade if it is among the highest 3 of your 5 R-Exercise scores across the semester.

The closing time for this R-Exercise is **4pm on Friday, October 13, 2023**. Please make sure that you have submitted your answers by this time. Remember that **you must click submit** before the deadline for your R-Exercise to be marked.

Please Note: If you encounter any technical issues with the R-Exercise, please email the CML coordinator at cml.2300@uq.edu.au. Do not email R-Exercise issues to the Course Coordinator or Course Administrator. Otherwise there may be a delay in responding to your enquiry.

SETutor will be available next Monday!

If you found these tutorials helpful, please answer the survey.

(If you didn't, please let me know how to improve them through the survey too ☺)

This is very valuable for us tutors!

https://eval.uq.edu.au/eus.onlinesurveyportal/Home/Survey?surveyid=768118861

- Download the files for tutorial 10 from Blackboard,
- save them into a folder for this tutorial.

Tutorial 10 [Week 11] Instrumental Variables Regression

Attached Files:
- Fertility_Description.pdf (65.732 KB)
- fertility.csv (4.729 MB)
- movies_description.pdf (95.789 KB)
- Movies.csv (97.014 KB)
- tutorial10.pdf (111.658 KB)

- Copy the code from Codeshare,
- https://codeshare.io/tut10
- Paste the code in a new script in RStudio,
- Save the script in the same folder as the data.

E12.1 How does fertility affect labor supply? That is, how much does a woman's labor supply fall when she has an additional child? In this exercise you will estimate this effect using data for married women from the 1980 U.S. Census.[1] The data are in the file `Fertility.csv` and described in `Fertility_Description.pdf`. The data set contains information on married women aged 21–35 with two or more children.

Variable	Description
morekids	=1 if mom had more than 2 children
boy1st	=1 if 1st child was a boy
boy2nd	=1 if 2nd child was a boy
samesex	=1 if 1st two children same sex
agem1	age of mom at census
black	=1 if mom is black
hispan	=1 if mom is Hispanic
othrace	=1 if mom is not black, Hispanic or white
weeksm1	mom's weeks worked in 1979

Documentation for Fertility and Fertility_Small Data Sets

These data are taken from the 1980 Census. These data were provided by Professor William Evans of the University of Maryland and were used in his paper with Joshua Angrist "Children and Thier Parents' Labor Supply: Evidence from Exogenous Variation in Family Size," American Economic Review, June 1998, Vol. 88, No. 3, 450-477. The file **Fertility.dta** (in STATA format) contains data on 254,654 women between the age of 21 and 35. The data in **Fertility** are a subset of the data used in the Angrist-Evans paper. (The file **Fertility_Small** contains data on a 30,000 randomly selected women from the **Fertility** data set. This smaller dataset is provided for students with memory limitations on their computer software.)

(a) Regress **weeksm1** on the indicator variable **morekids**, using OLS. On average, do women with more than two children work less than women with two children? How much less?

```
> OLS = lm_robust(weeksm1 ~ morekids, se_type = "stata")
> summary(OLS)

Call:
lm_robust(formula = weeksm1 ~ morekids, se_type = "stata")

Standard error type:  HC1

Coefficients:
            Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper     DF
(Intercept)   21.068    0.05607  375.76        0   20.959   21.178 254652
morekids      -5.387    0.08715  -61.81        0   -5.558   -5.216 254652

Multiple R-squared:  0.01431 ,  Adjusted R-squared:  0.0143
F-statistic:  3821 on 1 and 254652 DF,  p-value: < 2.2e-16
```

Table 1: Fertility and Labor Supply

	(1) OLS	(2) TSLS	(3) TSLS
(Intercept)	21.068***	21.421***	−4.792***
	(0.056)	(0.487)	(0.407)
morekids	−5.387***	−6.314***	−5.821***
	(0.087)	(1.275)	(1.246)
agem1			0.832***
			(0.023)
black			11.623***
			(0.229)
hispan			0.404
			(0.260)
othrace			2.131***
			(0.206)
R^2	0.014	0.014	0.044
Adj. R^2	0.014	0.014	0.044
Num. obs.	254654	254654	254654
RMSE	21.710	21.715	21.385

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

The coefficient is -5.39, which indicates that women with more than 2 children work 5.39 fewer weeks per year than women with 2 or fewer children.

(b) Explain why the OLS regression estimated in (a) is inappropriate for estimating the causal effect of fertility (`morekids`) on labor supply (`weeksm1`).

Both fertility and weeks worked are choice variables. A woman with a positive labor sup- ply regression error (a woman who works more than average) may also be a woman who is less likely to have an additional child. This would imply that `morekids` is correlated with the error, so that the OLS estimator of $\beta_{morekids}$ is biased.

(c) The data set contains the variable **samesex**, which is equal to 1 if the first two children are of the same sex (boy–boy or girl–girl) and equal to 0 otherwise. Are couples whose first two children are of the same sex more likely to have a third child? Is the effect large? Is it statistically significant?

```
> OLS.first1 = lm_robust(morekids ~ samesex, se_type = "stata")
> summary(OLS.first1)

Call:
lm_robust(formula = morekids ~ samesex, se_type = "stata")

Standard error type:  HC1

Coefficients:
            Estimate Std. Error t value   Pr(>|t|) CI Lower CI Upper     DF
(Intercept)  0.34642   0.001341  258.34  0.000e+00  0.34380  0.34905 254652
samesex      0.06753   0.001919   35.19 1.388e-270  0.06376  0.07129 254652

Multiple R-squared:  0.004835 , Adjusted R-squared:  0.004831
F-statistic:  1238 on 1 and 254652 DF,  p-value: < 2.2e-16
```

Table 2: First Stage Estimation of TSLS

	(2) TSLS	(3) TSLS
(Intercept)	0.346***	−0.140***
	(0.001)	(0.008)
samesex	0.068***	0.068***
	(0.002)	(0.002)
agem1		0.015***
		(0.000)
black		0.101***
		(0.004)
hispan		0.151***
		(0.004)
othrace		0.028***
		(0.005)
R^2	0.005	0.024
Adj. R^2	0.005	0.024
Num. obs.	254654	254654
F statistic	1238.171	1303.930
RMSE	0.484	0.480

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

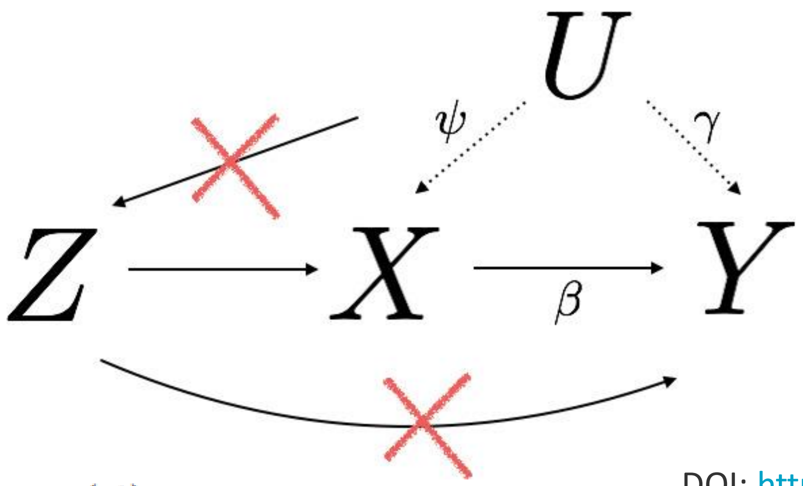
The linear regression of **morekids** on **samesex** (a linear probability model) yields

$$\widehat{\texttt{morekids}} = \underset{(0.001)}{0.346} + \underset{(0.002)}{0.068} \times \texttt{samesex}$$

so that couples with **samesex** = 1 are 6.8% more likely to have an additional child that couples with **samesex** = 0. The effect is highly significant (t-statistic = 35.2).

(d) Explain why `samesex` is a valid instrument for the instrumental variable regression of `weeksm1` on `morekids`.

(e) Is `samesex` a weak instrument?



DOI: https://doi.org/10.1145/3178876.3186151

Table 2: First Stage Estimation of TSLS

	(2) TSLS	(3) TSLS
(Intercept)	0.346***	−0.140***
	(0.001)	(0.008)
samesex	0.068***	0.068***
	(0.002)	(0.002)
agem1		0.015***
		(0.000)
black		0.101***
		(0.004)
hispan		0.151***
		(0.004)
othrace		0.028***
		(0.005)
R^2	0.005	0.024
Adj. R^2	0.005	0.024
Num. obs.	254654	254654
F statistic	1238.171	1303.930
RMSE	0.484	0.480

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

(d)

`samesex` is random and is unrelated to any of the other variables in the model including the error term in the labor supply equation. Thus, the instrument is exogenous. From (c), the first stage F-statistic is large ($F = 1238.2 > 10$) so the instrument is relevant. Together, these imply that `samesex` is a valid instrument.

(e)

No, see the answer to (d).

(f) Estimate the regression of `weeksm1` on `morekids`, using `samesex` as an instrument. How large is the fertility effect on labor supply?

```
> TSLS1 = ivreg(weeksm1 ~ morekids | samesex)
> summary(TSLS1)

Call:
ivreg(formula = weeksm1 ~ morekids | samesex)

Residuals:
   Min     1Q Median     3Q    Max 
-21.42 -21.42 -13.42  24.89  36.89 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   21.421      0.487  43.988  < 2e-16 ***
morekids      -6.314      1.275  -4.953  7.3e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 21.71 on 254652 degrees of freedom
Multiple R-Squared: 0.01388,	Adjusted R-squared: 0.01388 
Wald test: 24.54 on 1 and 254652 DF,  p-value: 7.296e-07
```

Table 1: Fertility and Labor Supply

	(1) OLS	(2) TSLS	(3) TSLS
(Intercept)	21.068^{***}	21.421^{***}	-4.792^{***}
	(0.056)	(0.487)	(0.407)
morekids	-5.387^{***}	-6.314^{***}	-5.821^{***}
	(0.087)	(1.275)	(1.246)
agem1			0.832^{***}
			(0.023)
black			11.623^{***}
			(0.229)
hispan			0.404
			(0.260)
othrace			2.131^{***}
			(0.206)
R^2	0.014	0.014	0.044
Adj. R^2	0.014	0.014	0.044
Num. obs.	254654	254654	254654
RMSE	21.710	21.715	21.385

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

See column (2) of Table 1. The estimated value of $\beta_{morekids} = -6.314$.

(f) Estimate the regression of **weeksm1** on **morekids**, using **samesex** as an instrument. How large is the fertility effect on labor supply?

(we can do it manually –
coefficients are right, but not standard error)

```
> # Extra Material:
> FS_morekids <- OLS.first1$coefficients[1] + OLS.first1$coefficients[2]*samesex
> TSLS1_hand <- lm_robust(weeksm1 ~ FS_morekids, se_type = "stata")
> summary(TSLS1_hand)

Call:
lm_robust(formula = weeksm1 ~ FS_morekids, se_type = "stata")

Standard error type:  HC1

Coefficients:
            Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper     DF
(Intercept)   21.421     0.4906  43.663 0.000e+00   20.460   22.383 254652
FS_morekids   -6.314     1.2836  -4.919 8.708e-07   -8.829   -3.798 254652

Multiple R-squared:  9.502e-05 ,        Adjusted R-squared:  9.109e-05
F-statistic:  24.2 on 1 and 254652 DF,  p-value: 8.708e-07
```

Table 1: Fertility and Labor Supply

	(1) OLS	(2) TSLS	(3) TSLS
(Intercept)	21.068^{***}	21.421^{***}	-4.792^{***}
	(0.056)	(0.487)	(0.407)
morekids	-5.387^{***}	-6.314^{***}	-5.821^{***}
	(0.087)	(1.275)	(1.246)
agem1			0.832^{***}
			(0.023)
black			11.623^{***}
			(0.229)
hispan			0.404
			(0.260)
othrace			2.131^{***}
			(0.206)
R^2	0.014	0.014	0.044
Adj. R^2	0.014	0.014	0.044
Num. obs.	254654	254654	254654
RMSE	21.710	21.715	21.385

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

(g) Do the results change when you include the variables **agem1**, **black**, **hispan**, and **othrace** in the labor supply regression (treating these variable as exogenous)? Explain why or why not.

```
> TSLS2 = ivreg(weeksm1 ~ morekids + agem1 + black + hispan + othrace |
+                  samesex + agem1 + black + hispan + othrace)
> summary(TSLS2)

Call:
ivreg(formula = weeksm1 ~ morekids + agem1 + black + hispan +
    othrace | samesex + agem1 + black + hispan + othrace)

Residuals:
   Min      1Q Median      3Q     Max
-36.34 -17.66 -10.99  22.72  45.15

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.79189    0.40657 -11.786   <2e-16 ***
morekids    -5.82105    1.24631  -4.671    3e-06 ***
agem1        0.83160    0.02289  36.336   <2e-16 ***
black       11.62327    0.22893  50.772   <2e-16 ***
hispan       0.40418    0.25986   1.555     0.12
othrace      2.13096    0.20586  10.352   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 21.38 on 254648 degrees of freedom
Multiple R-Squared: 0.04368,    Adjusted R-squared: 0.04366
Wald test:  1335 on 5 and 254648 DF,  p-value: < 2.2e-16
```

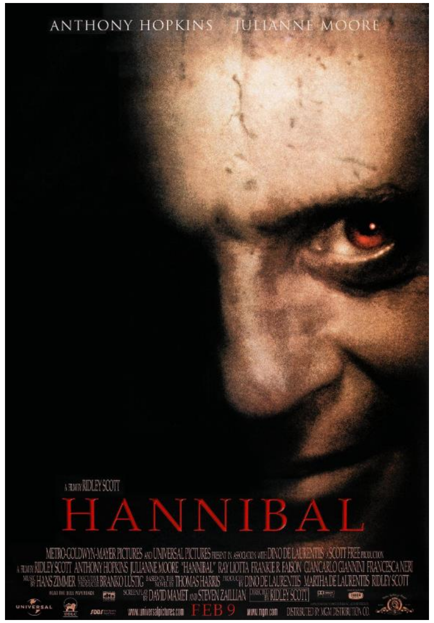
Table 1: Fertility and Labor Supply

	(1) OLS	(2) TSLS	(3) TSLS
(Intercept)	21.068***	21.421***	−4.792***
	(0.056)	(0.487)	(0.407)
morekids	−5.387***	−6.314***	−5.821***
	(0.087)	(1.275)	(1.246)
agem1			0.832***
			(0.023)
black			11.623***
			(0.229)
hispan			0.404
			(0.260)
othrace			2.131***
			(0.206)
R^2	0.014	0.014	0.044
Adj. R^2	0.014	0.014	0.044
Num. obs.	254654	254654	254654
RMSE	21.710	21.715	21.385

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

See column (3) of Table 1. The results do not change in an important way. The reason is that **samesex** is unrelated to **agem1**, **black**, **hispan**, and **othrace**, so that there is no omitted variable bias in IV regression in (f).

E12.2 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 that attracts many viewers. Alternatively, movie viewing may substitute for other activities (such as alcohol consumption) that lead to violent behavior, so that assaults should fall when more viewers are attracted to the cinema. Find the data file `Movies.csv`, which contains data on the number of assaults and movie attendance for 516 weekends from 1995 through 2004.[2] A detailed description is given in `Movies_Description.pdf`. The dataset includes weekend U.S. attendance for strongly violent movies (such as Hannibal), mildly violent movies (such as Spider-Man), and nonviolent movies (such as Finding Nemo). The dataset also includes a count of the number of assaults for the same weekend in a subset of counties in the United States. Finally, the dataset includes indicators for year, month, whether the weekend is a holiday, and various measures of the weather.



THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Variable Name	Description
Assaults and Movie Attendance	
assaults	number of assaults and intimidation in a subset of U.S. counties
attend_v	attendance stongly violent movies (in millions)
attend_m	attendance mildly violent movies (in millions)
attend_n	attendance nonviolent movies (in millions)
Weather, Holiday and Calendar Variables	
year1 to year10	indicator variable for year of the sample (1995-2004)
month1 to month12	indicator variables for month of the year (January-December)
h_chris	indicator variable for Christmas weekend
h_newyr	indicator variable for New Years weekend
h_easter	indicator variable for Easter weekend
h_july4	indicator variable for July 4 (U.S. Independence Day) weekend
h_mem	indicator variable for Memorial Day weekend
h_labor	indicator variable for Labor Day weekend
w_rain	fraction of locations with rain
w_snow	fraction of locations with snow
w_maxa	fraction of locations with maximum daily temperature between 80^{o}F and 90^{o}F
w_maxb	fraction of locations with maximum daily temperature between 90^{o}F and 100^{o}F
w_maxc	fraction of locations with maximum daily temperature greater than 100^{o}F
w_mina	fraction of locations with minimum daily temperature between 10^{o}F
w_minb	fraction of locations with minimum daily temperature less than 10^{o}F and 20^{o}F
w_minc	fraction of locations with minimum daily temperature less than 10^{o}F and 20^{o}F
Instruments	
pr_attend_v	predicted attendance violent movies
pr_attend_m	predicted attendance moderately violent movies
pr_attend_n	predicted attendance nonviolent movies
attend_v_f	attendance violent movies one week in the future
attend_m_f	attendance moderately violent movies one week in the future
attend_n_f	attendance nonviolent movies one week in the future
attend_v_b	attendance violent movies one week in the past
attend_m_b	attendance moderately violent movies one week in the past
attend_n_b	attendance nonviolent movies one week in the past

(a) i. Regress the logarithm of the number of assaults (**ln_assaults** = ln(**assaults**)) on the year and month indicators. Is there evidence of seasonality in assaults? That is, do there tend to be more assaults in some months than others? Explain.

```
> reg1 = lm_robust(ln_assaults ~ year2 + year3 + year4 + year5 + year6 + year7 +
+                    year8 + year9 + year10 + month2 + month3 + month4 + month5 +
+                    month6 + month7 + month8 + month9 + month10 + month11 +
+                    month12, se_type = "stata")
> linearHypothesis(reg1, c("month2=0", "month3=0", "month4=0", "month5=0",
+                          "month6=0", "month7=0", "month8=0", "month9=0",
+                          "month10=0", "month11=0", "month12=0"), test=c("F"))
Linear hypothesis test

Hypothesis:
month2 = 0
month3 = 0
month4 = 0
month5 = 0
month6 = 0
month7 = 0
month8 = 0
month9 = 0
month10 = 0
month11 = 0
month12 = 0

Model 1: restricted model
Model 2: ln_assaults ~ year2 + year3 + year4 + year5 + year6 + year7 +
    year8 + year9 + year10 + month2 + month3 + month4 + month5 +
    month6 + month7 + month8 + month9 + month10 + month11 + month12

  Res.Df Df      F    Pr(>F)
1    506
2    495 11 78.278 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The F-statistic on the 11 monthly indicators is 78.28 with a p-value that is essentially 0. Thus, there is strong evidence of seasonality in assaults. (The estimates imply that there are more assaults in the summer than in the winter.)

ii. Regress total movie attendance ($\mathtt{attend} = \mathtt{attend\_v} + \mathtt{attend\_m} + \mathtt{attend\_n}$) on the year and month indicators. Is there evidence of seasonality in movie attendance? Explain.

```
> reg2 = lm_robust(attend ~ year2 + year3 + year4 + year5 + year6 + year7 +
+                    year8 + year9 + year10 + month2 + month3 + month4 + month5 +
+                    month6 + month7 + month8 + month9 + month10 + month11 +
+                    month12, se_type = "stata")
> linearHypothesis(reg2, c("month2=0", "month3=0", "month4=0", "month5=0",
+                          "month6=0", "month7=0", "month8=0", "month9=0",
+                          "month10=0", "month11=0", "month12=0"), test=c("F"))
Linear hypothesis test

Hypothesis:
month2 = 0
month3 = 0
month4 = 0
month5 = 0
month6 = 0
month7 = 0
month8 = 0
month9 = 0
month10 = 0
month11 = 0
month12 = 0

Model 1: restricted model
Model 2: attend ~ year2 + year3 + year4 + year5 + year6 + year7 + year8 +
    year9 + year10 + month2 + month3 + month4 + month5 + month6 +
    month7 + month8 + month9 + month10 + month11 + month12

  Res.Df Df     F    Pr(>F)
1    506
2    495 11 58.57 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The F-statistic on the 11 monthly indicators is 58.57 with p-value that is essentially 0. Thus, there is strong evidence of seasonality in movie attendance. (The estimates imply that attendance is high in the summer.)

(b) Regress **ln_assaults** on **attend_v**, **attend_m**, **attend_n**, the year and month indicators, and the weather and holiday control variables available in the data set.

i. Based on the regression, does viewing a strongly violent movie increase or decrease assaults? By how much? Is the estimated effect statistically significant?

```
Call:
lm_robust(formula = ln_assaults ~ attend_v + attend_m + attend_n +
    year2 + year3 + year4 + year5 + year6 + year7 + year8 + year9 +
    year10 + month2 + month3 + month4 + month5 + month6 + month7 +
    month8 + month9 + month10 + month11 + month12 + h_chris +
    h_newyr + h_easter + h_july4 + h_mem + h_labor + w_maxa +

Standard error type:  HC1

Coefficients:
             Estimate Std. Error  t value   Pr(>|t|)  CI Lower   CI Upper  DF
(Intercept)  6.914316  0.0199267 346.9867  0.000e+00  6.875161  6.953e+00 478
attend_v    -0.003169  0.0010095  -3.1392  1.799e-03 -0.005152 -1.185e-03 478
attend_m    -0.003138  0.0007002  -4.4817  9.278e-06 -0.004514 -1.762e-03 478
attend_n    -0.002105  0.0007404  -2.8434  4.655e-03 -0.003560 -6.504e-04 478
year2        0.700811  0.0098018  71.4979 2.158e-257  0.681551  7.201e-01 478
year3        1.017069  0.0100452 101.2490  0.000e+00  0.997331  1.037e+00 478
year4        1.226729  0.0098888 124.0523  0.000e+00  1.207299  1.246e+00 478
year5        1.389127  0.0113609 122.2721  0.000e+00  1.366803  1.411e+00 478
year6        1.688294  0.0091901 183.7084  0.000e+00  1.670236  1.706e+00 478
year7        1.839547  0.0097684 188.3151  0.000e+00  1.820352  1.859e+00 478
year8        1.898393  0.0104275 182.0561  0.000e+00  1.877904  1.919e+00 478
year9        1.950122  0.0104520 186.5794  0.000e+00  1.929585  1.971e+00 478
year10       2.072147  0.0105116 197.1299  0.000e+00  2.051492  2.093e+00 478
month2      -0.007258  0.0124247  -0.5842  5.594e-01 -0.031672  1.716e-02 478
month3       0.012555  0.0124661   1.0071  3.144e-01 -0.011940  3.705e-02 478
month4       0.001864  0.0153559   0.1214  9.034e-01 -0.028309  3.204e-02 478
month5       0.007949  0.0158838   0.5004  6.170e-01 -0.023262  3.916e-02 478
month6      -0.029030  0.0167490  -1.7332  8.370e-02 -0.061940  3.881e-03 478
month7      -0.034348  0.0174474  -1.9687  4.957e-02 -0.068631 -6.541e-05 478
month8      -0.038458  0.0173885  -2.2117  2.746e-02 -0.072626 -4.291e-03 478
month9      -0.012855  0.0180179  -0.7135  4.759e-01 -0.048259  2.255e-02 478
month10     -0.002473  0.0166149  -0.1489  8.817e-01 -0.035120  3.017e-02 478
month11     -0.043193  0.0134709  -3.2064  1.434e-03 -0.069663 -1.672e-02 478
month12     -0.030492  0.0172486  -1.7678  7.773e-02 -0.064384  3.400e-03 478
h_chris     -0.087940  0.0343803  -2.5579  1.084e-02 -0.155495 -2.038e-02 478
h_newyr      0.245256  0.1112937   2.2037  2.802e-02  0.026571  4.639e-01 478
h_easter    -0.036940  0.0138321  -2.6706  7.830e-03 -0.064119 -9.761e-03 478
h_july4      0.035178  0.0148879   2.3628  1.853e-02  0.005924  6.443e-02 478
h_mem        0.005925  0.0117228   0.5054  6.135e-01 -0.017109  2.896e-02 478
h_labor      0.024149  0.0108193   2.2320  2.607e-02  0.002890  4.541e-02 478
w_maxa       0.109906  0.0111757   9.8343  6.590e-21  0.087946  1.319e-01 478
w_maxb       0.110748  0.0138328   8.0062  9.069e-15  0.083568  1.379e-01 478
w_maxc       0.042323  0.0440855   0.9600  3.375e-01 -0.044302  1.289e-01 478
w_mina      -0.340520  0.0576230  -5.9094  6.531e-09 -0.453746 -2.273e-01 478
w_minb      -0.172549  0.0375872  -4.5906  5.657e-06 -0.246405 -9.869e-02 478
w_minc      -0.119609  0.0184330  -6.4888  2.166e-10 -0.155828 -8.339e-02 478
w_rain      -0.032289  0.0120337  -2.6832  7.545e-03 -0.055934 -8.643e-03 478
w_snow      -0.061194  0.0422130  -1.4497  1.478e-01 -0.144140  2.175e-02 478

Multiple R-squared:  0.9959 ,   Adjusted R-squared:  0.9956
F-statistic:  3166 on 37 and 478 DF,  p-value: < 2.2e-16
```

The results are shown in the column labeled OLS in Table 3. An increase in strongly violent movie attendance of one million viewers is predicted to reduce assaults by 0.32%. The coefficient is statistically significant at the 1% significance level.

ii. Does attendance at strongly violent movies affect assaults differently than attendance at moderately violent movies? Differently than attendance at nonviolent movies?

```
Hypothesis:
attend_v - attend_m = 0

Model 1: restricted model
Model 2: ln_assaults ~ attend_v + attend_m + attend_n + year2 + year3 +
    year4 + year5 + year6 + year7 + year8 + year9 + year10 +
    month2 + month3 + month4 + month5 + month6 + month7 + month8 +
    month9 + month10 + month11 + month12 + h_chris + h_newyr +
    h_easter + h_july4 + h_mem + h_labor + w_maxa + w_maxb +
    w_maxc + w_mina + w_minb + w_minc + w_rain + w_snow

  Res.Df Df      F Pr(>F)
1    479
2    478  1 0.0014 0.9698
```

```
Hypothesis:
attend_v - attend_n = 0

Model 1: restricted model
Model 2: ln_assaults ~ attend_v + attend_m + attend_n + year2 + year3 +
    year4 + year5 + year6 + year7 + year8 + year9 + year10 +
    month2 + month3 + month4 + month5 + month6 + month7 + month8 +
    month9 + month10 + month11 + month12 + h_chris + h_newyr +
    h_easter + h_july4 + h_mem + h_labor + w_maxa + w_maxb +
    w_maxc + w_mina + w_minb + w_minc + w_rain + w_snow

  Res.Df Df      F Pr(>F)
1    479
2    478  1 1.6318 0.2021
```

```
Hypothesis:
attend_v - attend_m = 0
attend_v - attend_n = 0

Model 1: restricted model
Model 2: ln_assaults ~ attend_v + attend_m + attend_n + year2 + year3 +
    year4 + year5 + year6 + year7 + year8 + year9 + year10 +
    month2 + month3 + month4 + month5 + month6 + month7 + month8 +
    month9 + month10 + month11 + month12 + h_chris + h_newyr +
    h_easter + h_july4 + h_mem + h_labor + w_maxa + w_maxb +
    w_maxc + w_mina + w_minb + w_minc + w_rain + w_snow

  Res.Df Df      F Pr(>F)
1    480
2    478  2 1.5471 0.2139
```

The F-statistic suggests that the coefficients β_v, β_m, and β_n are not statistically significantly different from one another.

iii. A strongly violent blockbuster movie is released, and the weekend's attendance at strongly violent movies increases by 6 million; meanwhile, attendance falls by 2 million for moderately violent movies and by 1 million for nonviolent movies. What is the predicted effect on assaults? Construct a 95% confidence interval for the change in assaults. [*Hint*: Review Section 7.3 and material surrounding Equations (8.7) and (8.8).]

```
> confint(glht(OLS, linfct = c("6*attend_v - 2*attend_m - attend_n = 0")))

	 Simultaneous Confidence Intervals

Fit: lm_robust(formula = ln_assaults ~ attend_v + attend_m + attend_n + 
    year2 + year3 + year4 + year5 + year6 + year7 + year8 + year9 + 
    year10 + month2 + month3 + month4 + month5 + month6 + month7 + 
    month8 + month9 + month10 + month11 + month12 + h_chris + 
    h_newyr + h_easter + h_july4 + h_mem + h_labor + w_maxa + 
    w_maxb + w_maxc + w_mina + w_minb + w_minc + w_rain + w_snow, 
    se_type = "stata")

Quantile = 1.96
95% family-wise confidence level
 

Linear Hypotheses:
                                           Estimate   lwr        upr       
6 * attend_v - 2 * attend_m - attend_n == 0 -0.0106321 -0.0204376 -0.0008266
```

The question asks for an estimate and standard error for $6\beta_v - 2\beta_m - \beta_n$. The OLS estimate for this coefficient is -0.011. It shows a decrease in assaults of 1.1%. The 95% confidence interval is -0.020 to -0.0008 (or -2.0% to -0.08%).

(c) It is difficult to control for all the variables that affect assaults and that might be correlated with movie attendance. For example, the effect of the weather on assaults and movie attendance is only crudely approximated by the weather variables in the data set. However, the data set does include a set of instruments, **pr_attend_v**, **pr_attend_m**, and **pr_attend_n**, that are correlated with attendance but are (arguably) uncorrelated with weekend-specific factors (such as the weather) that affect both assaults and movie attendance. These instruments use historical attendance patterns, not information on a particular weekend, to predict a film's attendance in a given weekend. For example, if a film's attendance is high in the second week of its release, then this can be used to predict that its attendance was also high in the first week of its release. (The details of the construction of these instruments are available in the Dahl and DellaVigna paper referenced in footnote 5.) Run the regression from part (b) (including year, month, holiday, and weather controls) but now using **pr_ attend_v**, **pr_attend_m**, and **pr_attend_n** as instruments for **attend_v**, **attend_m**, and **attend_n**. Use this regression to answer (b)i–(b)iii.

```
Call:
ivreg(formula = ln_assaults ~ attend_v + attend_m + attend_n +
    year2 + year3 + year4 + year5 + year6 + year7 + year8 + year9 +
    year10 + month2 + month3 + month4 + month5 + month6 + month7 +
    month8 + month9 + month10 + month11 + month12 + h_chris +
    h_newyr + h_easter + h_july4 + h_mem + h_labor + w_maxa +
    w_maxb + w_maxc + w_mina + w_minb + w_minc + w_rain + w_snow |
    pr_attend_v + pr_attend_m + pr_attend_n + year2 + year3 +
        year4 + year5 + year6 + year7 + year8 + year9 + year10 +
        month2 + month3 + month4 + month5 + month6 + month7 +
        month8 + month9 + month10 + month11 + month12 + h_chris +
        h_newyr + h_easter + h_july4 + h_mem + h_labor + w_maxa +
        w_maxb + w_maxc + w_mina + w_minb + w_minc + w_rain +
        w_snow)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.33227 -0.02473 -0.00252  0.02318  0.18905 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  6.9241781  0.0182776 378.834  < 2e-16 ***
attend_v    -0.0038738  0.0011131  -3.480 0.000547 ***
attend_m    -0.0038930  0.0007748  -5.025 7.14e-07 ***
attend_n    -0.0027221  0.0007892  -3.449 0.000612 ***
year2        0.7019494  0.0085867  81.748  < 2e-16 ***
year3        1.0191737  0.0087824 116.047  < 2e-16 ***
year4        1.2293360  0.0089676 137.086  < 2e-16 ***
year5        1.3910170  0.0087758 158.506  < 2e-16 ***
year6        1.6901854  0.0086467 195.473  < 2e-16 ***
year7        1.8422489  0.0089949 204.810  < 2e-16 ***
year8        1.9022983  0.0094491 201.320  < 2e-16 ***
year9        1.9534999  0.0090464 215.943  < 2e-16 ***
year10       2.0750249  0.0091914 225.758  < 2e-16 ***
month2      -0.0077646  0.0095867  -0.810 0.418380    
month3       0.0110923  0.0101777   1.090 0.276320    
month4      -0.0005341  0.0125280  -0.043 0.966013    
month5       0.0076986  0.0146063   0.527 0.598386    
month6      -0.0269358  0.0157359  -1.712 0.087591 .  
month7      -0.0307414  0.0179397  -1.714 0.087251 .  
month8      -0.0378368  0.0168999  -2.239 0.025623 *  
month9      -0.0162870  0.0153787  -1.059 0.290106    
month10     -0.0044822  0.0126541  -0.354 0.723341    
month11     -0.0408120  0.0109236  -3.736 0.000209 ***
month12     -0.0309746  0.0099068  -3.127 0.001876 ** 
h_chris     -0.0839750  0.0236690  -3.548 0.000427 ***
h_newyr      0.2510780  0.0232545  10.797  < 2e-16 ***
h_easter    -0.0357587  0.0146070  -2.448 0.014722 *  
h_july4      0.0348679  0.0203022   1.717 0.086546 .  
h_mem        0.0112309  0.0152665   0.736 0.462301    
h_labor      0.0235949  0.0142537   1.655 0.098510 .  
w_maxa       0.1101143  0.0134986   8.157 3.05e-15 ***
w_maxb       0.1123803  0.0186237   6.034 3.20e-09 ***
w_maxc       0.0469039  0.0699786   0.670 0.503015    
w_mina      -0.3440012  0.0397046  -8.664  < 2e-16 ***
w_minb      -0.1735184  0.0270351  -6.418 3.32e-10 ***
w_minc      -0.1178151  0.0168387  -6.997 8.90e-12 ***
w_rain      -0.0316781  0.0128550  -2.464 0.014080 *  
w_snow      -0.0599756  0.0298137  -2.012 0.044814 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.04196 on 478 degrees of freedom
Multiple R-Squared: 0.9959,	Adjusted R-squared: 0.9956 
Wald test:  3130 on 37 and 478 DF,  p-value: < 2.2e-16
```

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

```
> linearHypothesis(TSLS1, c("attend_v=attend_m"))
Linear hypothesis test

Hypothesis:
attend_v - attend_m = 0

Model 1: restricted model
Model 2: ln_assaults ~ attend_v + attend_m + attend_n + year2 + year3 +
    year4 + year5 + year6 + year7 + year8 + year9 + year10 +
    month2 + month3 + month4 + month5 + month6 + month7 + month8 +
    month9 + month10 + month11 + month12 + h_chris + h_newyr +
    h_easter + h_july4 + h_mem + h_labor + w_maxa + w_maxb +
    w_maxc + w_mina + w_minb + w_minc + w_rain + w_snow | pr_attend_v +
    pr_attend_m + pr_attend_n + year2 + year3 + year4 + year5 +
    year6 + year7 + year8 + year9 + year10 + month2 + month3 +
    month4 + month5 + month6 + month7 + month8 + month9 + month10 +
    month11 + month12 + h_chris + h_newyr + h_easter + h_july4 +
    h_mem + h_labor + w_maxa + w_maxb + w_maxc + w_mina + w_minb +
    w_minc + w_rain + w_snow

  Res.Df Df Chisq Pr(>Chisq)
1    479
2    478  1 4e-04     0.9838
```

```
> linearHypothesis(TSLS1, c("attend_v=attend_n"))
Linear hypothesis test

Hypothesis:
attend_v - attend_n = 0

Model 1: restricted model
Model 2: ln_assaults ~ attend_v + attend_m + attend_n + year2 + year3 +
    year4 + year5 + year6 + year7 + year8 + year9 + year10 +
    month2 + month3 + month4 + month5 + month6 + month7 + month8 +
    month9 + month10 + month11 + month12 + h_chris + h_newyr +
    h_easter + h_july4 + h_mem + h_labor + w_maxa + w_maxb +
    w_maxc + w_mina + w_minb + w_minc + w_rain + w_snow | pr_attend_v +
    pr_attend_m + pr_attend_n + year2 + year3 + year4 + year5 +
    year6 + year7 + year8 + year9 + year10 + month2 + month3 +
    month4 + month5 + month6 + month7 + month8 + month9 + month10 +
    month11 + month12 + h_chris + h_newyr + h_easter + h_july4 +
    h_mem + h_labor + w_maxa + w_maxb + w_maxc + w_mina + w_minb +
    w_minc + w_rain + w_snow

  Res.Df Df  Chisq Pr(>Chisq)
1    479
2    478  1 1.3437     0.2464
```

```
> linearHypothesis(TSLS1, c("attend_v=attend_m", "attend_v=attend_n"))
Linear hypothesis test

Hypothesis:
attend_v - attend_m = 0
attend_v - attend_n = 0

Model 1: restricted model
Model 2: ln_assaults ~ attend_v + attend_m + attend_n + year2 + year3 +
    year4 + year5 + year6 + year7 + year8 + year9 + year10 +
    month2 + month3 + month4 + month5 + month6 + month7 + month8 +
    month9 + month10 + month11 + month12 + h_chris + h_newyr +
    h_easter + h_july4 + h_mem + h_labor + w_maxa + w_maxb +
    w_maxc + w_mina + w_minb + w_minc + w_rain + w_snow | pr_attend_v +
    pr_attend_m + pr_attend_n + year2 + year3 + year4 + year5 +
    year6 + year7 + year8 + year9 + year10 + month2 + month3 +
    month4 + month5 + month6 + month7 + month8 + month9 + month10 +
    month11 + month12 + h_chris + h_newyr + h_easter + h_july4 +
    h_mem + h_labor + w_maxa + w_maxb + w_maxc + w_mina + w_minb +
    w_minc + w_rain + w_snow

  Res.Df Df  Chisq Pr(>Chisq)
1    480
2    478  2 3.1674     0.2052
```

```
> confint(glht(TSLS1, linfct = c("6*attend_v - 2*attend_m - attend_n = 0")))

         Simultaneous Confidence Intervals

Fit: ivreg(formula = ln_assaults ~ attend_v + attend_m + attend_n +
    year2 + year3 + year4 + year5 + year6 + year7 + year8 + year9 +
    year10 + month2 + month3 + month4 + month5 + month6 + month7 +
    month8 + month9 + month10 + month11 + month12 + h_chris +
    h_newyr + h_easter + h_july4 + h_mem + h_labor + w_maxa +
    w_maxb + w_maxc + w_mina + w_minb + w_minc + w_rain + w_snow |
    pr_attend_v + pr_attend_m + pr_attend_n + year2 + year3 +
        year4 + year5 + year6 + year7 + year8 + year9 + year10 +
        month2 + month3 + month4 + month5 + month6 + month7 +
        month8 + month9 + month10 + month11 + month12 + h_chris +
        h_newyr + h_easter + h_july4 + h_mem + h_labor + w_maxa +
        w_maxb + w_maxc + w_mina + w_minb + w_minc + w_rain +
        w_snow)

Quantile = 1.96
95% family-wise confidence level


Linear Hypotheses:
                                           Estimate  lwr       upr
6 * attend_v - 2 * attend_m - attend_n == 0 -0.012735 -0.023918 -0.001551
```

Table 3: Violent Movie and Violent Behavior

	(1) OLS	(2) IV	(3) TSLS
(Intercept)	6.9143***	6.9242***	6.9225***
	(0.0199)	(0.0183)	(0.0264)
attend_v	−0.0032**	−0.0039***	−0.0032
	(0.0010)	(0.0011)	(0.0021)
attend_m	−0.0031***	−0.0039***	−0.0041**
	(0.0007)	(0.0008)	(0.0015)
attend_n	−0.0021**	−0.0027***	−0.0026
	(0.0007)	(0.0008)	(0.0015)
R^2	0.9959	0.9959	0.9959
Adj. R^2	0.9956	0.9956	0.9956
Num. obs.	516	516	516
RMSE	0.0419	0.0420	0.0420

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

i The results are shown in the column labeled IV in Table 3. An increase in strongly violent movie attendance of one million viewers is predicted to reduce assaults by 0.39%. The coefficient is statistically significant at the 1% significance level.

ii The F-statistic suggests that the coefficients β_v, β_m, and β_n are not statistically significantly different from one another.

iii The TSLS estimate for this coefficient is -0.013. It shows a decrease in assaults of 1.3%. The 95% confidence interval is -0.024 to -0.0016 (or -2.4% to -0.16%).

(d) The intuition underlying the instruments in (c) is that attendance in a given week is correlated with attendance in surrounding weeks. For each move category, the data set includes attendance in surrounding weeks. Run the regression using the instruments **attend_v_f**, **attend_m_f**, **attend_n_f**, **attend_v_b**, **attend_m_b**, and **attend_n_b** instead of the instruments used in part (c). Use this regression to answer (b)i–(b)iii.

```
Call:
ivreg(formula = ln_assaults ~ attend_v + attend_m + attend_n +
    year2 + year3 + year4 + year5 + year6 + year7 + year8 + year9 +
    year10 + month2 + month3 + month4 + month5 + month6 + month7 +
    month8 + month9 + month10 + month11 + month12 + h_chris +
    h_newyr + h_easter + h_july4 + h_mem + h_labor + w_maxa +
    w_maxb + w_maxc + w_mina + w_minb + w_minc + w_rain + w_snow |
    attend_v_f + attend_m_f + attend_n_f + attend_v_b + attend_m_b +
        attend_n_b + year2 + year3 + year4 + year5 + year6 +
        year7 + year8 + year9 + year10 + month2 + month3 + month4 +
        month5 + month6 + month7 + month8 + month9 + month10 +
        month11 + month12 + h_chris + h_newyr + h_easter + h_july4 +
        h_mem + h_labor + w_maxa + w_maxb + w_maxc + w_mina +
        w_minb + w_minc + w_rain + w_snow)

Residuals:
      Min        1Q    Median        3Q       Max
-0.332642 -0.023980 -0.002817  0.023012  0.191329

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.9225164  0.0264364 261.855  < 2e-16 ***
attend_v    -0.0031738  0.0020536  -1.545 0.122898
attend_m    -0.0041215  0.0015128  -2.725 0.006676 **
attend_n    -0.0025823  0.0015398  -1.677 0.094187 .
year2        0.7024852  0.0088294  79.562  < 2e-16 ***
year3        1.0194733  0.0095465 106.790  < 2e-16 ***
year4        1.2299978  0.0100482 122.410  < 2e-16 ***
year5        1.3912278  0.0093842 148.252  < 2e-16 ***
year6        1.6905674  0.0092587 182.593  < 2e-16 ***
year7        1.8436000  0.0101817 181.070  < 2e-16 ***
year8        1.9044430  0.0116528 163.433  < 2e-16 ***
year9        1.9536683  0.0108640 179.830  < 2e-16 ***
year10       2.0750073  0.0105644 196.416  < 2e-16 ***
month2      -0.0083817  0.0096658  -0.867 0.386295
month3       0.0107342  0.0105718   1.015 0.310448
month4      -0.0005969  0.0133165  -0.045 0.964265
month5       0.0093665  0.0151267   0.619 0.536079
month6      -0.0261280  0.0162121  -1.612 0.107703
month7      -0.0309198  0.0191857  -1.612 0.107709
month8      -0.0378970  0.0170794  -2.219 0.026964 *
month9      -0.0162274  0.0168159  -0.965 0.335029
month10     -0.0051668  0.0132325  -0.390 0.696372
month11     -0.0419820  0.0121161  -3.465 0.000578 ***
month12     -0.0309227  0.0099586  -3.105 0.002015 **
h_chris     -0.0846492  0.0249752  -3.389 0.000759 ***
h_newyr      0.2506722  0.0258369   9.702  < 2e-16 ***
h_easter    -0.0352876  0.0148117  -2.382 0.017590 *
h_july4      0.0359201  0.0204176   1.759 0.079171 .
h_mem        0.0115023  0.0179979   0.639 0.523071
h_labor      0.0237070  0.0143132   1.656 0.098317 .
w_maxa       0.1100234  0.0135237   8.136 3.57e-15 ***
w_maxb       0.1115675  0.0191333   5.831 1.02e-08 ***
w_maxc       0.0467240  0.0710200   0.658 0.510919
w_mina      -0.3460864  0.0404670  -8.552  < 2e-16 ***
w_minb      -0.1730187  0.0271446  -6.374 4.34e-10 ***
w_minc      -0.1177419  0.0171635  -6.860 2.14e-11 ***
w_rain      -0.0323418  0.0129912  -2.490 0.013131 *
w_snow      -0.0593807  0.0300969  -1.973 0.049073 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.04201 on 478 degrees of freedom
Multiple R-Squared: 0.9959,     Adjusted R-squared: 0.9956
Wald test:  3121 on 37 and 478 DF,  p-value: < 2.2e-16
```

```
> linearHypothesis(TSLS2, c("attend_v=attend_m"))
Linear hypothesis test

Hypothesis:
attend_v - attend_m = 0

Model 1: restricted model
Model 2: ln_assaults ~ attend_v + attend_m + attend_n + year2 + year3 +
    year4 + year5 + year6 + year7 + year8 + year9 + year10 +
    month2 + month3 + month4 + month5 + month6 + month7 + month8 +
    month9 + month10 + month11 + month12 + h_chris + h_newyr +
    h_easter + h_july4 + h_mem + h_labor + w_maxa + w_maxb +
    w_maxc + w_mina + w_minb + w_minc + w_rain + w_snow | attend_v_f +
    attend_m_f + attend_n_f + attend_v_b + attend_m_b + attend_n_b +
    year2 + year3 + year4 + year5 + year6 + year7 + year8 + year9 +
    year10 + month2 + month3 + month4 + month5 + month6 + month7 +
    month8 + month9 + month10 + month11 + month12 + h_chris +
    h_newyr + h_easter + h_july4 + h_mem + h_labor + w_maxa +
    w_maxb + w_maxc + w_mina + w_minb + w_minc + w_rain + w_snow

  Res.Df Df  Chisq Pr(>Chisq)
1    479
2    478  1 0.3799     0.5376
> linearHypothesis(TSLS2, c("attend_v=attend_n"))
Linear hypothesis test

Hypothesis:
attend_v - attend_n = 0

Model 1: restricted model
Model 2: ln_assaults ~ attend_v + attend_m + attend_n + year2 + year3 +
    year4 + year5 + year6 + year7 + year8 + year9 + year10 +
    month2 + month3 + month4 + month5 + month6 + month7 + month8 +
    month9 + month10 + month11 + month12 + h_chris + h_newyr +
    h_easter + h_july4 + h_mem + h_labor + w_maxa + w_maxb +
    w_maxc + w_mina + w_minb + w_minc + w_rain + w_snow | attend_v_f +
    attend_m_f + attend_n_f + attend_v_b + attend_m_b + attend_n_b +
    year2 + year3 + year4 + year5 + year6 + year7 + year8 + year9 +
    year10 + month2 + month3 + month4 + month5 + month6 + month7 +
    month8 + month9 + month10 + month11 + month12 + h_chris +
    h_newyr + h_easter + h_july4 + h_mem + h_labor + w_maxa +
    w_maxb + w_maxc + w_mina + w_minb + w_minc + w_rain + w_snow

  Res.Df Df  Chisq Pr(>Chisq)
1    479
2    478  1 0.1752     0.6756
```

```
> linearHypothesis(TSLS2, c("attend_v=attend_m", "attend_v=attend_n"))
Linear hypothesis test

Hypothesis:
attend_v - attend_m = 0
attend_v - attend_n = 0

Model 1: restricted model
Model 2: ln_assaults ~ attend_v + attend_m + attend_n + year2 + year3 +
    year4 + year5 + year6 + year7 + year8 + year9 + year10 +
    month2 + month3 + month4 + month5 + month6 + month7 + month8 +
    month9 + month10 + month11 + month12 + h_chris + h_newyr +
    h_easter + h_july4 + h_mem + h_labor + w_maxa + w_maxb +
    w_maxc + w_mina + w_minb + w_minc + w_rain + w_snow | attend_v_f +
    attend_m_f + attend_n_f + attend_v_b + attend_m_b + attend_n_b +
    year2 + year3 + year4 + year5 + year6 + year7 + year8 + year9 +
    year10 + month2 + month3 + month4 + month5 + month6 + month7 +
    month8 + month9 + month10 + month11 + month12 + h_chris +
    h_newyr + h_easter + h_july4 + h_mem + h_labor + w_maxa +
    w_maxb + w_maxc + w_mina + w_minb + w_minc + w_rain + w_snow

  Res.Df Df  Chisq Pr(>Chisq)
1    480
2    478  2 1.9047     0.3858
```

```
> confint(glht(TSLS2, linfct = c("6*attend_v - 2*attend_m - attend_n = 0")))

         Simultaneous Confidence Intervals

Fit: ivreg(formula = ln_assaults ~ attend_v + attend_m + attend_n +
    year2 + year3 + year4 + year5 + year6 + year7 + year8 + year9 +
    year10 + month2 + month3 + month4 + month5 + month6 + month7 +
    month8 + month9 + month10 + month11 + month12 + h_chris +
    h_newyr + h_easter + h_july4 + h_mem + h_labor + w_maxa +
    w_maxb + w_maxc + w_mina + w_minb + w_minc + w_rain + w_snow |
    attend_v_f + attend_m_f + attend_n_f + attend_v_b + attend_m_b +
        attend_n_b + year2 + year3 + year4 + year5 + year6 +
        year7 + year8 + year9 + year10 + month2 + month3 + month4 +
        month5 + month6 + month7 + month8 + month9 + month10 +
        month11 + month12 + h_chris + h_newyr + h_easter + h_july4 +
        h_mem + h_labor + w_maxa + w_maxb + w_maxc + w_mina +
        w_minb + w_minc + w_rain + w_snow)

Quantile = 1.96
95% family-wise confidence level


Linear Hypotheses:
                                       Estimate  lwr       upr
6 * attend_v - 2 * attend_m - attend_n == 0 -0.008218 -0.027112 0.010676
```

Table 3: Violent Movie and Violent Behavior

	(1) OLS	(2) IV	(3) TSLS
(Intercept)	6.9143***	6.9242***	6.9225***
	(0.0199)	(0.0183)	(0.0264)
attend_v	−0.0032**	−0.0039***	−0.0032
	(0.0010)	(0.0011)	(0.0021)
attend_m	−0.0031***	−0.0039***	−0.0041**
	(0.0007)	(0.0008)	(0.0015)
attend_n	−0.0021**	−0.0027***	−0.0026
	(0.0007)	(0.0008)	(0.0015)
R^2	0.9959	0.9959	0.9959
Adj. R^2	0.9956	0.9956	0.9956
Num. obs.	516	516	516
RMSE	0.0419	0.0420	0.0420

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

i The results are shown in the column labeled TSLS in Table 3. An increase in strongly violent movie attendance of one million viewers is predicted to reduce assaults by 0.32%. The coefficient is not statistically significant at the 10% significance level.

ii The F-statistic suggests that the coefficients β_v, β_m, and β_n are not statistically significantly different from one another.

iii The TSLS estimate for this coefficient is -0.008. It shows a decrease in assaults of 0.8%. The 95% confidence interval is -0.027 to 0.011 (or -2.7% to 1.1%).

(e) There are nine instruments listed in (c) and (d), but only three are needed for identification. Carry out the test for over-identification summarized in Key Concept 12.6. What do you conclude about the validity of the instruments?

```
Call:
ivreg(formula = ln_assaults ~ attend_v + attend_m + attend_n +
    year2 + year3 + year4 + year5 + year6 + year7 + year8 + year9 +
    year10 + month2 + month3 + month4 + month5 + month6 + month7 +
    month8 + month9 + month10 + month11 + month12 + h_chris +
    h_newyr + h_easter + h_july4 + h_mem + h_labor + w_maxa +
    w_maxb + w_maxc + w_mina + w_minb + w_minc + w_rain + w_snow |
    pr_attend_v + pr_attend_m + pr_attend_n + attend_v_f + attend_m_f +
        attend_n_f + attend_v_b + attend_m_b + attend_n_b + year2 +
        year3 + year4 + year5 + year6 + year7 + year8 + year9 +
        year10 + month2 + month3 + month4 + month5 + month6 +
        month7 + month8 + month9 + month10 + month11 + month12 +
        h_chris + h_newyr + h_easter + h_july4 + h_mem + h_labor +
        w_maxa + w_maxb + w_maxc + w_mina + w_minb + w_minc +
        w_rain + w_snow)

Residuals:
      Min        1Q    Median        3Q       Max
-0.332697 -0.024633 -0.002676  0.023207  0.188275

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.9263773  0.0182239 380.070  < 2e-16 ***
attend_v    -0.0039808  0.0011089  -3.590 0.000365 ***
attend_m    -0.0040052  0.0007700  -5.202 2.93e-07 ***
attend_n    -0.0028900  0.0007817  -3.697 0.000243 ***
year2        0.7020895  0.0085914  81.720  < 2e-16 ***
year3        1.0194320  0.0087848 116.045  < 2e-16 ***
year4        1.2298091  0.0089666 137.155  < 2e-16 ***
year5        1.3913327  0.0087781 158.501  < 2e-16 ***
year6        1.6904496  0.0086493 195.443  < 2e-16 ***
year7        1.8425995  0.0089958 204.829  < 2e-16 ***
year8        1.9029120  0.0094434 201.508  < 2e-16 ***
year9        1.9541341  0.0090401 216.164  < 2e-16 ***
year10       2.0755821  0.0091882 225.896  < 2e-16 ***
month2      -0.0078968  0.0095925  -0.823 0.410790
month3       0.0107183  0.0101811   1.053 0.292981
month4      -0.0010738  0.0125305  -0.086 0.931743
month5       0.0073361  0.0146141   0.502 0.615908
month6      -0.0267383  0.0157445  -1.698 0.090110 .
month7      -0.0303921  0.0179466  -1.693 0.091017 .
month8      -0.0379641  0.0169102  -2.245 0.025222 *
month9      -0.0171513  0.0153775  -1.115 0.265259
month10     -0.0049876  0.0126576  -0.394 0.693726
month11     -0.0402322  0.0109233  -3.683 0.000257 ***
month12     -0.0310432  0.0099130  -3.132 0.001845 **
h_chris     -0.0830251  0.0236752  -3.507 0.000496 ***
h_newyr      0.2524376  0.0232505  10.857  < 2e-16 ***
h_easter    -0.0354392  0.0146149  -2.425 0.015683 *
h_july4      0.0347194  0.0203151   1.709 0.088091 .
h_mem        0.0122783  0.0152575   0.805 0.421371
h_labor      0.0234736  0.0142626   1.646 0.100460
w_maxa       0.1101330  0.0135073   8.154 3.13e-15 ***
w_maxb       0.1129294  0.0186324   6.061 2.75e-09 ***
w_maxc       0.0485348  0.0700157   0.693 0.488521
w_mina      -0.3440790  0.0397289  -8.661  < 2e-16 ***
w_minb      -0.1736866  0.0270522  -6.420 3.28e-10 ***
w_minc      -0.1175959  0.0168483  -6.980 9.93e-12 ***
w_rain      -0.0314974  0.0128627  -2.449 0.014694 *
w_snow      -0.0601316  0.0298323  -2.016 0.044395 *

Diagnostic tests:
                         df1 df2 statistic p-value
Weak instruments (attend_v)   9 472   802.134  <2e-16 ***
Weak instruments (attend_m)   9 472   503.362  <2e-16 ***
Weak instruments (attend_n)   9 472   356.038  <2e-16 ***
Wu-Hausman                    3 475     1.747   0.157
Sargan                        6  NA     9.227   0.161
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.04198 on 478 degrees of freedom
Multiple R-Squared: 0.9959,	Adjusted R-squared: 0.9956
Wald test:  3126 on 37 and 478 DF,  p-value: < 2.2e-16
```

(e) There are nine instruments listed in (c) and (d), but only three are needed for identification. Carry out the test for over-identification summarized in Key Concept 12.6. What do you conclude about the validity of the instruments?

The J-statistic is 9.23, which is distributed χ^2_6 under the null hypothesis that the additional instruments are exogenous. As the p-value is 0.16, we do not reject the null hypothesis at the 10% level.

Sargan–Hansen test

From Wikipedia, the free encyclopedia

The **Sargan–Hansen test** or **Sargan's J test** is a statistical test used for testing over-identifying restrictions in a statistical model. It was proposed by John Denis Sargan in 1958,[1] and several variants were derived by him in 1975.[2] Lars Peter Hansen re-worked through the derivations and showed that it can be extended to general non-linear GMM in a time series context.[3]

The Sargan test is based on the assumption that model parameters are identified via a priori restrictions on the coefficients, and tests the validity of over-identifying restrictions. The test statistic can be computed from residuals from instrumental variables regression by constructing a quadratic form based on the cross-product of the residuals and exogenous variables.[4]: 132–33 Under the null hypothesis that the over-identifying restrictions are valid, the statistic is asymptotically distributed as a chi-square variable with $(m-k)$ degrees of freedom (where m is the number of instruments and k is the number of endogenous variables).

(f) Based on your analysis, what do you conclude about the effect of violent movies on (short-run) violent behavior?

Movie attendance appears to reduce assaults, but there is little evidence of a differential effect of violent movies. This result is consistent with a mechanism in which movies attendance is a substitute for other activities, such as drinking, that increase assaults.

Thank you

Francisco Tavares Garcia

Academic Tutor | School of Economics

tavaresgarcia.github.io

Reference

Stock, J. H., & Watson, M. W. (2019). Introduction to Econometrics, Global Edition, 4th edition. Pearson Education Limited.