Instrument Variable EE12.2

EE12.2

我們直接從課本的 Empirical Exercise 12.2 來複習如何用 Stata 處理內生性問題，並實作 2SLS。

EE12.2 Questions

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. On the text website, **http://www.pearsonglobaleditions.com**, you will find the data file **Movies**, which contains data on the number of assaults and movie attendance for 516 weekends from 1995 through 2004.[7] A detailed description is given in **Movies_Description**, available on the website. The data set 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 data set also includes a count of the number of assaults for the same weekend in a subset of counties in the United States. Finally, the data set includes indicators for year, month, whether the weekend is a holiday, and various measures of the weather.

EE12.2 Questions

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.

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

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?

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

[7]These are aggregated versions of data provided by Gordon Dahl of University of California–San Diego and Stefano DellaVigna of University of California–Berkeley and were used in their paper "Does Movie

EE12.2 Questions

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).]

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 (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 IV regression to answer (b)(i)–(b)(iii).

EE12.2 Questions

d. The intuition underlying the instruments in (c) is that attendance in a given week is correlated with attendance in surrounding weeks. For each movie 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 (c). Use this IV regression to answer (b)(i)–(b)(iii).

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

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

EE12.2 Data Descrptions

Movie Data

1. Observations: 516 weekends
2. Time Period : 1995-2004

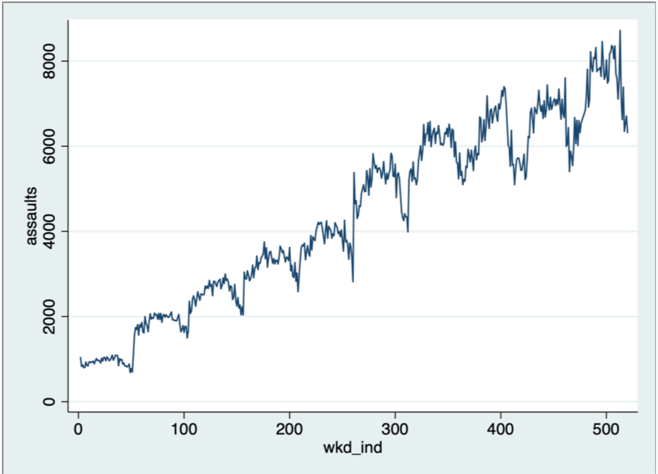
Variable Name	Description
	Assaults and Movie Attendance
assaults	number of assaults and intimidation in a subset of U.S. counties
attend_v	attendance stongly violent movies (in millions)
attend_m	attendance mildly violent movies (in millions)
attend_n	attendance nonviolent movies (in millions)
	Weather, 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 80ºF and 90ºF
w_maxb	fraction of locations with maximum daily temperature between 90ºF and 100ºF
w_maxc	fraction of locations with maximum daily temperature greater than 100ºF
w_mina	fraction of locations with minimum daily temperature less than 10ºF
w_minb	fraction of locations with minimum daily temperature between 10ºF and 20ºF
w_minc	fraction of locations with minimum daily temperature between 20ºF and 32ºF
	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

EE12.2 a(i.)

To detect whether there is time trend or not:

$$log(assaults) = \beta_0 + \psi_1 year + \psi_2 month + u$$

Or just simply graph a twoway plot with the time indicator.

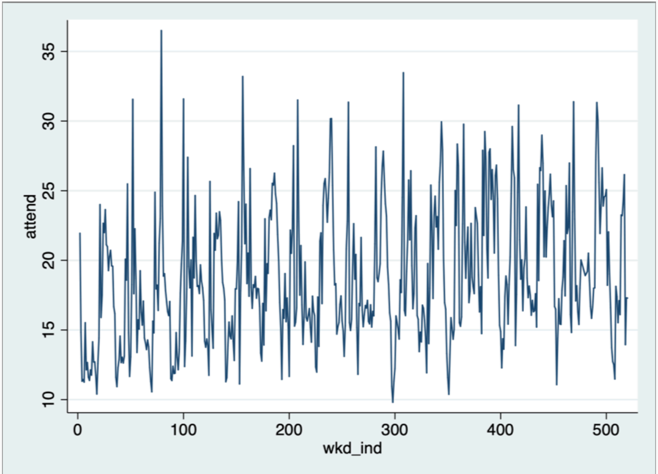


EE12.2 a(ii.)

To detect whether there is time trend or not:

$$attendance = \beta_0 + \phi_1 year + \phi_2 month + \tilde{u}$$

Or just simply graph a twoway plot with the time indicator.



EE12.2 b

The original model is:

$$log(assaults) = \beta_0 + \beta_1 attend\_v + \beta_2 attend\_m + \beta_3 attend\_n + \phi Time + \psi Controls$$

And the variables of interest are `attend_v`, `attend_m`, `attend_n`

Till now, we do not say the variables of interest are exog. or endog.

EE12.2 b(iii.)

The original model is:

$$log(assaults) = \beta_0 + \beta_1 attend\_v + \beta_2 attend\_m + \beta_3 attend\_n + \phi Time + \psi Controls$$

What we want to know is the change in the dependent variable: $\Delta assaults$.

Let's start from:

$$\Delta log(assaults) = \beta_1 \Delta attend\_v + \beta_2 \Delta attend\_m + \beta_3 \Delta attend\_n$$

Then,

$$\Delta log(\widehat{assaults}) = \hat{\beta}_1 \Delta attend\_v + \hat{\beta}_2 \Delta attend\_m + \hat{\beta}_3 \Delta attend\_n$$

With the description in the questions:

$$\Delta log(\widehat{assaults}) = 6\hat{\beta}_1 - 2\hat{\beta}_2 - \hat{\beta}_3$$

EE12.2 b(iii.)

$$\Delta \widehat{log(assaults)} = 6\hat{\beta}_1 - 2\hat{\beta}_2 - \hat{\beta}_3$$

And we can easily calculate the estimate: $\Delta \widehat{log(assaults)} = -.01063206$

That is: $\Delta \widehat{assaults} \approx e^{-.01063206}$

EE12.2 b(iii.) conti.

Given $\Delta \widehat{log(assaults)} = 6\hat{\beta}_1 - 2\hat{\beta}_2 - \hat{\beta}_3$

To obtain $se(\Delta \widehat{log(assaults)}) = se(6\hat{\beta}_1 - 2\hat{\beta}_2 - \hat{\beta}_3)$,

we need to apply approaches in section 7.3.

We may simplify the model by:

$$y = \beta_0 + \beta_1 v + \beta_2 m + \beta_3 n + u$$

Now let's focus on $\beta_1 v + \beta_2 m + \beta_3 n$ only.

Our goal is to have a explanatory variable which has a coefficient equals to the above number.

EE12.2 b(iii.) conti.

Simplified model:

$$y = \beta_0 + \beta_1 v + \beta_2 m + \beta_3 n + u$$

$$y = \beta_1 v + \beta_2 m + \beta_3 n - 6\beta_1 n + 2\beta_2 n + 6\beta_1 n - 2\beta_2 n + u$$

$$y = \beta_1(v + 6n) + \beta_2(m - 2n) + (\beta_3 - 6\beta_1 + 2\beta_2)n + u$$

Now we can simply use OLS to obtain the s.e.

EE12.2 c

Now we want to use IVs.

Notations:

$Y: dependentvariable$

$X: possiblyendog.expanatoryvariables$

$W: exog.expanatoryvariables$

$Z: instrumentvariables$

Recall the 2SLS:

- Regress X on Z, W
- Obtain $\hat{X}$
- Regress Y on $\hat{X}, W$
- Obtain the coef. of $\hat{X}$

EE12.2 c

Denote

$Y: log(assaults)$

$X: attend\_v, attend\_m, attend\_n$

$W: Time, Holiday, Weather$

$Z: pr\_attend\_v, pr\_attend\_m, pr\_attend\_n$

which are predictions based on historical attendance patterns.

THINK: Why are these variable instruments?

We'll demonstrate the manual 2SLS approach first.

EE12.2 c

We may simply use `ivreg` or `ivregress`

The format would be:

```
ivreg Y (X=Z) W, r
```

or:

```
ivregress 2sls Y (X=Z) W, r
ivregress gmm Y (X=Z) W, r
```

EE12.2 d

Now change the IVs:

$$Z : attend\_v\_f, attend\_m\_f, attend\_n\_f,$$

attend_v_b, attend_m_b, attend_n_b

which are the lagged terms and the future terms.

EE12.2 e

For Overidentifying Restriction Test, we first apply J-Test in textbook p.449, then we'll demonstrate a simple and more general command in Stata.

Note that the model now is:

$Y = \beta_{2SLS}\hat{X} + \gamma W + u$

We may obtain $\hat{u}_{2SLS} = Y - \hat{\beta}_{2SLS}\hat{X} + \hat{\gamma} W$

Then we regress $\hat{u}_{2SLS}$ on Z, W, that is:

$$\hat{u}_{2SLS} = \delta Z + \tilde{\gamma} W + e$$

EE12.2 e

And we can obtain the J-statistic from $J = mF$ where

$m =$ the number of instruments

$F =$ the joint test statistic of $\delta = 0$

$k =$ the number of exog. variables

Under homosk., $J \sim \chi^2(m-k)$

EE12.2 e

An easy command is:

```
ivregress Y (X=Z) W, r
estat overid
```

will give you the test statistic under robust variance covariance estimator.