

7) Two-Stage Least Squares (2SLS)

Vitor Kamada

June 2018

Tables, Graphics, and Figures from

**Mastering 'Metrics: The Path from Cause
to Effect**

Angrist & Pischke (2014): Chapter 3.3

$$y = \beta_0 + \beta_1 x + u$$

$$\text{I) } \text{Cov}(z, u) = 0$$

$$\text{II) } \text{Cov}(z, x) \neq 0$$

$$\text{Cov}(z, y) = \beta_1 \text{Cov}(z, x) + \text{Cov}(z, u)$$

$$\hat{\beta}_1 = \frac{\text{Cov}(z, y)}{\text{Cov}(z, x)}$$

$$X = \pi_0 + \pi_1 Z + v$$

$$\pi_1 = \frac{\text{Cov}(z, x)}{\text{Var}(z)}$$

$$H_0 : \pi_1 = 0$$

Does High Fertility Rate Perpetuate Poverty?

Population Grows Faster than Food (Malthus)

Quantity-Quality Tradeoff (Gary Becker)

Negative correlation between average family size
and development indicators like education

China One Child Policy in 1979

Angrist, Lavy, and Schlosser (2010)

Second Birth	Children per Family
Singleton	3.6
Twin	3.92

1% of mothers have twins

Second-Born Sibling	Children per Family
Opposite Sex	3.6
Same Sex	3.68

Half of families have either two boys or two girls

Two-Stage Least Squares (2SLS)

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u_1$$

$$y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3 + v_2$$

$$\text{Cov}(z_i, v_2) = 0, \quad i = 1, 2, 3$$

$$\pi_2 \neq 0 \text{ or } \pi_3 \neq 0$$

$$y_1 \text{ on } \hat{y}_2 \text{ and } z_1$$

Quantity-Quality First Stages

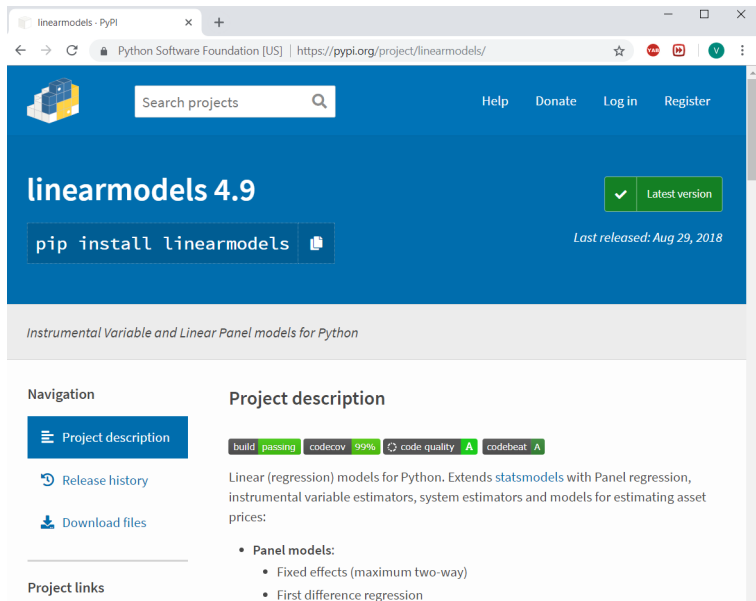
	Twins instruments		Same-sex instruments		Twins and same- sex instruments
	(1)	(2)	(3)	(4)	(5)
Second-born twins	.320 (.052)	.437 (.050)			.449 (.050)
Same-sex sibships			.079 (.012)	.073 (.010)	.076 (.010)
Male		-.018 (.010)		-.020 (.010)	-.020 (.010)
Controls	No	Yes	No	Yes	Yes

Sample Size is 89,445

Estimates of the Quantity-Quality Trade-off

Dependent variable	OLS estimates (1)	2SLS estimates		
		Twins instruments (2)	Same-sex instruments (3)	Twins and same- sex instruments (4)
Years of schooling	-.145 (.005)	.174 (.166)	.318 (.210)	.237 (.128)
High school graduate	-.029 (.001)	.030 (.028)	.001 (.033)	.017 (.021)
Some college (for age ≥ 24)	-.023 (.001)	.017 (.052)	.078 (.054)	.048 (.037)
College graduate (for age ≥ 24)	-.015 (.001)	-.021 (.045)	.125 (.053)	.052 (.032)

How to Install Packages in Anaconda?



The screenshot shows the PyPI project page for 'linearmodels'. The browser address bar shows the URL 'https://pypi.org/project/linearmodels/'. The page has a blue header with a search bar and navigation links: 'Help', 'Donate', 'Log in', and 'Register'. The main content area features the package name 'linearmodels 4.9' in large white text. To the right of the name is a green button with a checkmark and the text 'Latest version'. Below the name is a dark blue button with the text 'pip install linearmodels' and a small icon of a terminal. To the right of this button, it says 'Last released: Aug 29, 2018'. Below the main content area, there is a light gray section with the text 'Instrumental Variable and Linear Panel models for Python'. On the left side, there is a 'Navigation' section with a blue button for 'Project description' and links for 'Release history' and 'Download files'. On the right side, there is a 'Project description' section with a row of badges: 'build passing', 'codecov 99%', 'code quality A', and 'codebeat A'. Below the badges, the description reads: 'Linear (regression) models for Python. Extends statsmodels with Panel regression, instrumental variable estimators, system estimators and models for estimating asset prices:'. Below the description, there is a list of 'Panel models:' with two items: 'Fixed effects (maximum two-way)' and 'First difference regression'.

linearmodels 4.9

Search projects

Help Donate Log in Register

✓ Latest version

`pip install linearmodels`

Last released: Aug 29, 2018

Instrumental Variable and Linear Panel models for Python

Navigation

- Project description
- Release history
- Download files

Project links

Project description

build passing codecov 99% code quality A codebeat A

Linear (regression) models for Python. Extends statsmodels with Panel regression, instrumental variable estimators, system estimators and models for estimating asset prices:

- Panel models:
 - Fixed effects (maximum two-way)
 - First difference regression

Go to Anaconda Environments

Anaconda Navigator

File Help

ANACONDA NAVIGATOR

Home

Environments

Learning

Community

Search Environments

base (root) ▶

py27

py34

Installed

Channels

Update index...

Search Packages

Name	Description
✓ _ipyw_lab_nb_ext...	
✓ _r-mutex	
✓ alabaster	Configurable, python 2+3 compatible sphinx theme
✓ altair	
✓ anaconda-client	Anaconda.org command line client library

1) Click in the green arrow of “base (root)” to “Open Terminal”

2) Then, type: “pip install linearmodels”

"The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions", *Econometrica*, 55, 765-799.

```
import pandas as pd
```

```
import numpy as np
```

```
from linearmodels.iv import IV2SLS
```

```
df = pd.read_stata('C:\\Users\\Vitor\\Desktop\\ECO 6100  
Introduction to Econometrics (Fall 2018)\\Lectures\\7) Two-Stage  
Least Squares (2SLS)\\MROZ.dta')
```

```
df[['wage','educ','exper','motheduc','fatheduc']].describe()
```

	wage	educ	exper	motheduc	fatheduc
count	428.000000	753.000000	753.000000	753.000000	753.000000
mean	4.177680	12.286853	10.63081	9.250996	8.808765
std	3.310283	2.280246	8.06913	3.367468	3.572290
min	0.128200	5.000000	0.00000	0.000000	0.000000
25%	2.262600	12.000000	4.00000	7.000000	7.000000
50%	3.481900	12.000000	9.00000	10.000000	7.000000
75%	4.970750	13.000000	15.00000	12.000000	12.000000
max	25.000000	17.000000	45.00000	17.000000	17.000000

```
mod1 = 'np.log(wage) ~ 1 + exper + expersq + educ'
```

```
reg1 = IV2SLS.from_formula(mod1, df).fit(cov_type='robust')  
print(reg1)
```

	Parameter	Std. Err.	T-stat	P-value
Intercept	-0.5220	0.2007	-2.6010	0.0093
exper	0.0416	0.0152	2.7344	0.0063
expersq	-0.0008	0.0004	-1.9402	0.0524
educ	0.1075	0.0132	8.1697	0.0000

```
mod2 = 'np.log(wage) ~ 1 + exper + expersq +  
[educ ~ motheduc + fatheduc]'
```

```
reg2 = IV2SLS.from_formula(mod2, df).fit(cov_type='robust')  
print(reg2)
```

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
Intercept	0.0481	0.4278	0.1124	0.9105	-0.7903	0.8865
exper	0.0442	0.0155	2.8546	0.0043	0.0138	0.0745
expersq	-0.0009	0.0004	-2.1001	0.0357	-0.0017	-5.997e-05
educ	0.0614	0.0332	1.8503	0.0643	-0.0036	0.1264

```
=====  
Endogenous: educ  
Instruments: motheduc, fatheduc  
Robust Covariance (Heteroskedastic)
```

```
print(reg2.first_stage)
```

```
----- educ
R-squared                0.2115
Partial R-squared        0.2076
Shea's R-squared         0.2076
Partial F-statistic       100.22
P-value (Partial F-stat)  0.0000
Partial F-stat Distn      chi2(2)
=====
Intercept                9.1026
                        (21.588)
exper                    0.0452
                        (1.0854)
expersq                  -0.0010
                        (-0.7671)
motheduc                 0.1576
                        (4.4718)
fatheduc                 0.1895
                        (5.8771)
```

T-stats reported in parentheses