Module 2: Demand for Cigarettes and Instrumental Variables

Part 3: Application of IV to Demand Estimation

Ian McCarthy | Emory University
Econ 470 & HLTH 470

Naive estimate

Clearly a strong relationship between prices and sales. For example, just from OLS:

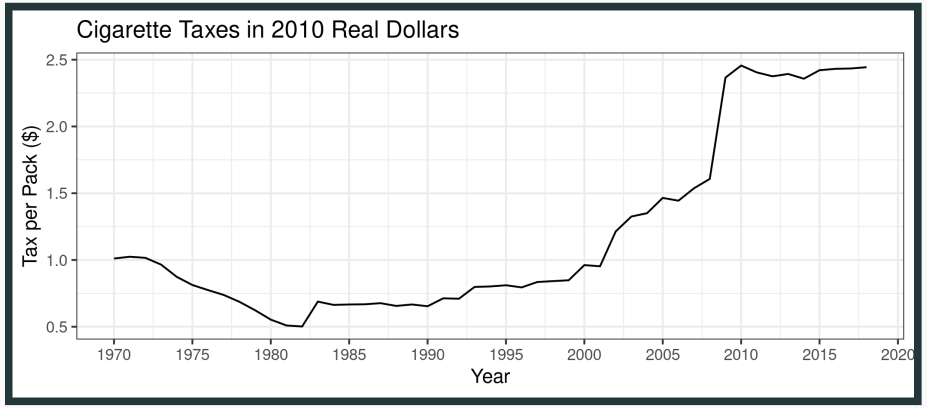
```
## 
## Call:
## lm(formula = ln_sales ~ ln_price, data = cig.data)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -1.23899 -0.17057  0.02239  0.18605  1.13866 
## 
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  4.689838   0.007209  650.55   <2e-16 ***
## ln_price    -0.420307   0.006464  -65.02   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.3073 on 2497 degrees of freedom
## Multiple R-squared:  0.6287,    Adjusted R-squared:  0.6285 
## F-statistic:  4228 on 1 and 2497 DF,  p-value: < 2.2e-16
```

Is this causal?

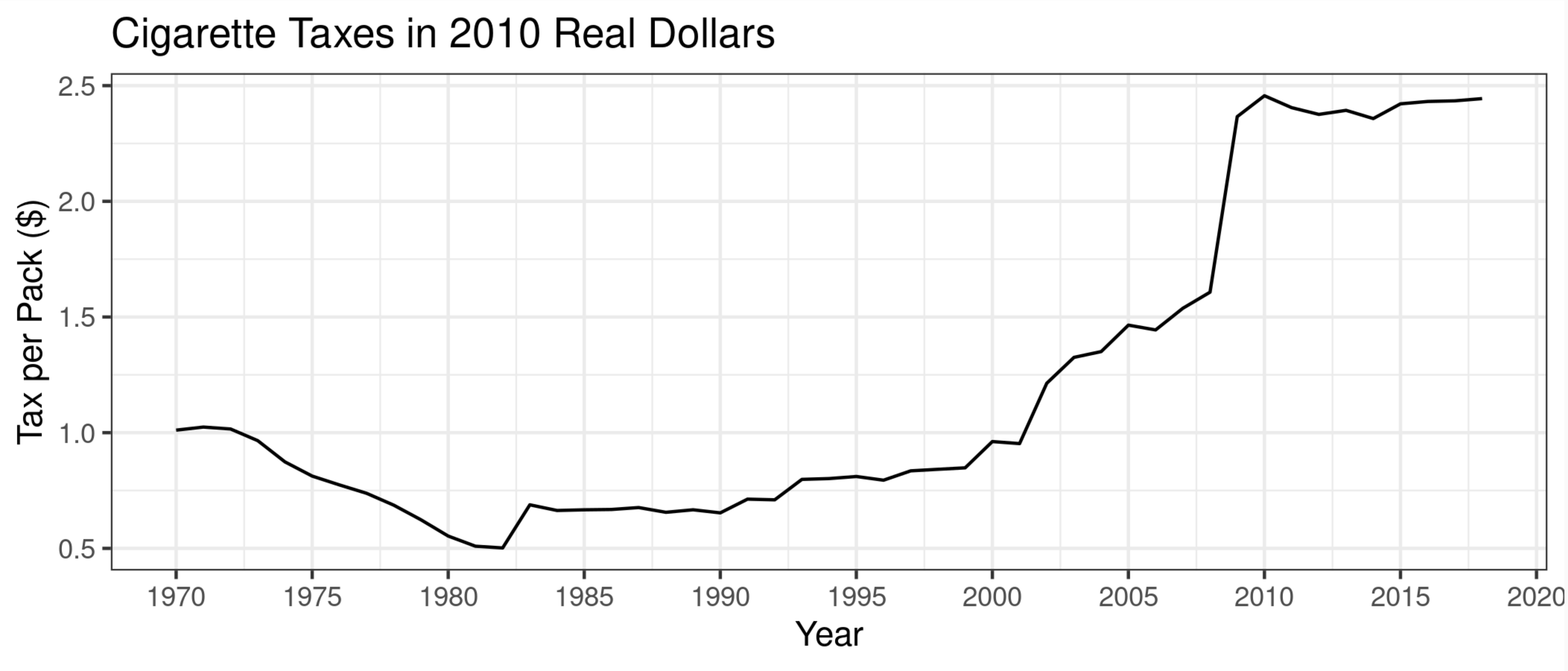
- But is that the true demand curve?
- Aren't other things changing that tend to reduce cigarette sales?

Tax as an IV

```
cig.data %>%
  ggplot(aes(x=Year,y=total_tax_cpi)) +
  stat_summary(fun.y="mean",geom="line") +
  labs(
    x="Year",
    y="Tax per Pack ($)",
    title="Cigarette Taxes in 2010 Real Dollars"
  ) + theme_bw() +
  scale_x_continuous(breaks=seq(1970, 2020, 5))
```



Tax as an IV


Cigarette Taxes in 2010 Real Dollars
Tax per Pack ($)
2.5
2.0
1.5
1.0
0.5
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
Year

IV Results

```
##
## Call:
## ivreg(formula = ln_sales ~ ln_price | total_tax_cpi, data = cig.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.24595 -0.23048  0.02863  0.23548  1.30999
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.805691   0.009703  495.29   <2e-16 ***
## ln_price    -0.619142   0.011128  -55.64   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3608 on 2497 degrees of freedom
## Multiple R-Squared: 0.488,    Adjusted R-squared: 0.4878
## Wald test:  3096 on 1 and 2497 DF,  p-value: < 2.2e-16
```

Two-stage equivalence

```r
step1 <- lm(ln_price ~ total_tax_cpi, data=cig.data)
pricehat <- predict(step1)
step2 <- lm(ln_sales ~ pricehat, data=cig.data)
summary(step2)
```

```
##
## Call:
## lm(formula = ln_sales ~ pricehat, data = cig.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.10960 -0.17805  0.01867  0.18697  1.14907
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.805691   0.008195  586.41   <2e-16 ***
## pricehat    -0.619142   0.009399  -65.87   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3048 on 2497 degrees of freedom
## Multiple R-squared:  0.6348,    Adjusted R-squared:  0.6346
## F-statistic:  4339 on 1 and 2497 DF,  p-value: < 2.2e-16
```

Different specifications

	Log Sales per Capita					
	OLS			IV		
	(1)	(2)	(3)	(4)	(5)	(6)
Log Price	-0.953***	-0.921***	-1.213***	-1.072***	-1.036***	-1.523***
	(0.012)	(0.008)	(0.034)	(0.014)	(0.010)	(0.041)
State FE	No	Yes	Yes	No	Yes	Yes
Year FE	No	No	Yes	No	No	Yes
Observations	2,499	2,499	2,499	2,499	2,499	2,499

Note:

Test the IV

	Log Price			Log Sales		
	First Stage			Reduced Form		
	(1)	(2)	(3)	(4)	(5)	(6)
Tax per Pack	0.444***	0.474***	0.187***	-0.476***	-0.491***	-0.284***
	(0.006)	(0.006)	(0.002)	(0.007)	(0.006)	(0.007)
State FE	No	Yes	Yes	No	Yes	Yes
Year FE	No	No	Yes	No	No	Yes
Observations	2,499	2,499	2,499	2,499	2,499	2,499

Note:

Summary

1. Most elasticities of around -0.25% to -0.37%

2. Much larger elasticities when including year fixed effects

3. Perhaps not too outlandish given more recent evidence: NBER Working Paper.

Some other IV issues

1. IV estimators are biased. Performance in finite samples is questionable.
2. IV estimators provide an estimate of a Local Average Treatment Effect (LATE), which is only the same as the ATT under some conditions or assumptions.
3. What about lots of instruments? The finite sample problem is more important and we may try other things (JIVE).

   The National Bureau of Economic Research (NBER) has a great resource here for understanding instruments in practice.

Quick IV Review

1. When do we consider IV as a potential identification strategy?

2. What are the main IV assumptions (and what do they mean)?

3. How do we test for those assumptions?

Review of IV and Homework 3 in R

Some coding pointers to keep in mind:

- Function for extracting top observations, `top_n` (negative values to look from the bottom up)
- Syntax for `ivreg`, part of `ivpack`
- Using `stargazer` for regression output

Other points from weekly surveys

- Examples of IV in real life
    - Month of birth
    - Vietnam lottery
    - Medicaid lottery in Oregon
- Problem with compliers
- Is a non-complier just in the control group?