ECON-C4100 - Capstone: Econometrics I

Lectures 10&11: Causal parameters part II - Instrumental variables regression

Otto Toivanen

Learning outcomes

- At the end of these lectures, you understand

1. what simultaneous causality means
2. what is meant by an **endogeneity** problem
3. why it causes bias in the parameters
4. what an instrumental variable is and why it solves the endogeneity problem
5. what characteristics are required of an instrumental variable
6. what one should pay attention to when using an instrumental variable

Learning outcomes

- At the end of these lectures, you understand

7. what a **reduced form** equation/parameter is
8. what a **structural** equation/parameter is
9. how to "manually" estimate a model with simultaneous causality
10. what **2SLS** estimation means, how you do it and why it is used

Overview

- Demand experiment, market data analysis.
- Simultaneous causality.
- IV regression and 2SLS.
- NOTE: Instrumental variables are used in a large variety of contexts.
- We are exploring it in a particular but historically and practically very important setting.
- In *Applied Microeconometrics I* and *II* you will learn more about IV, its use and the interpretation of results.

Simultaneous causality

- Your task is to estimate the demand function for a homogeneous good sold at unit price.
- How would you do this in an experiment?

Simultaneous causality

- Your task is to estimate the demand function for a homogeneous good sold at unit price.
- How would you do this in an experiment?
- By changing the price yourself (”at random”) and observing how many units are sold at each price.

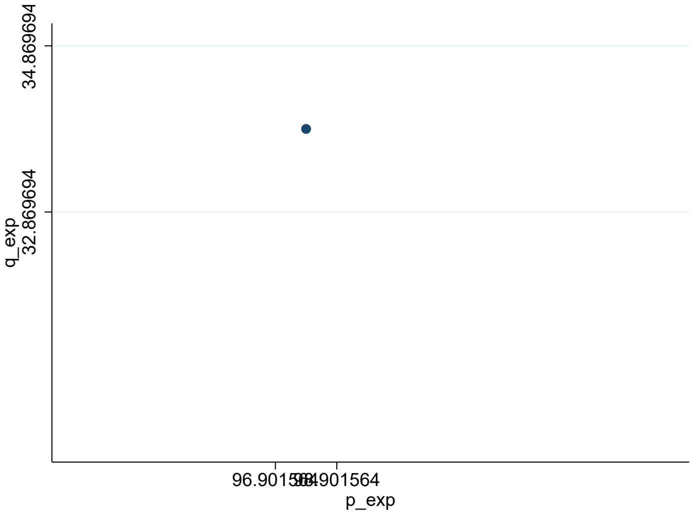
Experiment

- What does "choosing prices" at random mean?
  1. We offer different randomized prices to individual consumers
  2. We offer different randomized prices each to a group of consumers
  3. Think either of geographically separate markets, or a given market over time.

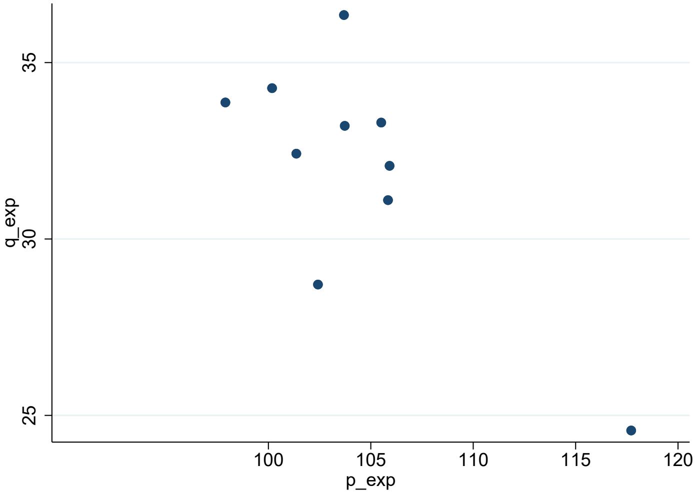
Experiment

- We record quantity sold at different prices
- We study the outcomes
- For illustration, I have conducted such an experiment in my computer.
- We will get to the details of how I do it later, but now just imagine I have conducted the experiment in a real market.

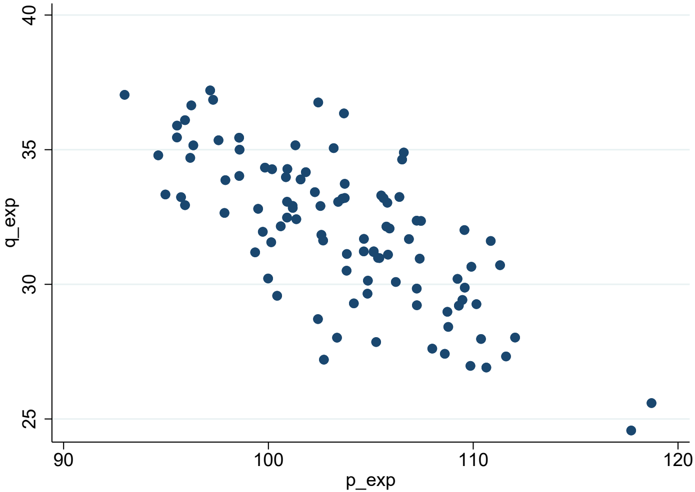
Experiment



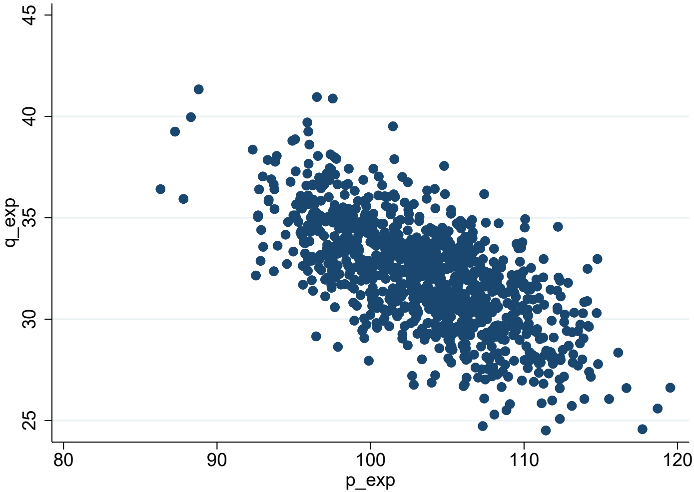
Experiment



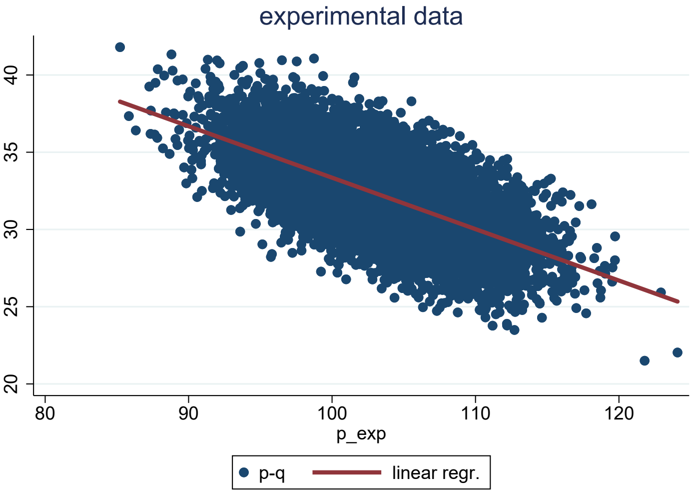
Experiment



Experiment



Experiment



experimental data
40
35
30
25
20
80
90
100
110
120
p_exp
p-q
linear regr.

Experiment

- Question: why does sold quantity vary between two experiments where the prices are identical?

Experiment

- Question: why does sold quantity vary between two experiments where the prices are identical?
- Answer: Demand is stochastic from the viewpoint of the econometrician.
- Let's study a simple set-up (the one I used in the experiment) in more detail.

Linear demand

- Demand function

$$Q_i = a - bP_i + \epsilon_i$$

- $a =$ average intercept.
- $b =$ slope.
- $\epsilon_i = =$ market specific deviation from the average intercept.
- $i =$ a particular market realisation.
- Question: where does this demand function come from?
- Answer: from consumers making utility-maximizing choices.
- Exercise: what does the utility function look like that produces a linear demand function?

Linear demand

- Inverse demand function

$$P_i = \frac{a}{b} - \frac{1}{b}Q_i + \frac{1}{b}\epsilon_i = \alpha + \beta Q_i + \tilde{\epsilon}_i$$

Regression analysis

```
. regr q_exp p_exp

      Source |       SS           df       MS      Number of obs   =    10,000
-------------+----------------------------------   F(1, 9998)      =   6935.10
       Model |  27500.3163         1  27500.3163   Prob > F        =    0.0000
    Residual |  39645.8875     9,998  3.96538183   R-squared       =    0.4096
-------------+----------------------------------   Adj R-squared   =    0.4095
       Total |  67146.2038     9,999  6.71529191   Root MSE        =    1.9913

------------------------------------------------------------------------------
       q_exp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       p_exp |  -.3329958   .0039986   -83.28   0.000    -.3408339   -.3251577
       _cons |   66.65341   .4145481   160.79   0.000     65.84081      67.466
------------------------------------------------------------------------------
```

Robustness analysis

```
. gen p_exp2             = p_exp^2

. regr q_exp p_exp*

      Source |       SS           df       MS      Number of obs   =    10,000
-------------+----------------------------------   F(2, 9997)      =   3468.48
       Model |  27506.3195         2  13753.1598   Prob > F        =    0.0000
    Residual |  39639.8843     9,997  3.96517798   R-squared       =    0.4096
-------------+----------------------------------   Adj R-squared   =    0.4095
       Total |  67146.2038     9,999  6.71529191   Root MSE        =    1.9913

------------------------------------------------------------------------------
       q_exp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       p_exp |  -.4806486   .1200668    -4.00   0.000    -.7160038   -.2452935
      p_exp2 |   .0007134   .0005798     1.23   0.219    -.0004231    .0018498
       _cons |   74.27605   6.208917    11.96   0.000     62.10532    86.44677
------------------------------------------------------------------------------
```

Parameters for inverse demand function

```
. scalar alpha_exp        = - _b[_cons] / _b[p_exp]

. scalar beta_exp         = -1 / _b[p_exp]

. scalar list alpha_exp beta_exp
 alpha_exp =  200.16291
  beta_exp =   3.003041
```

Market outcomes

- Assume you are an outside observer of a market say a prospective buyer of a firm or the competition authority.

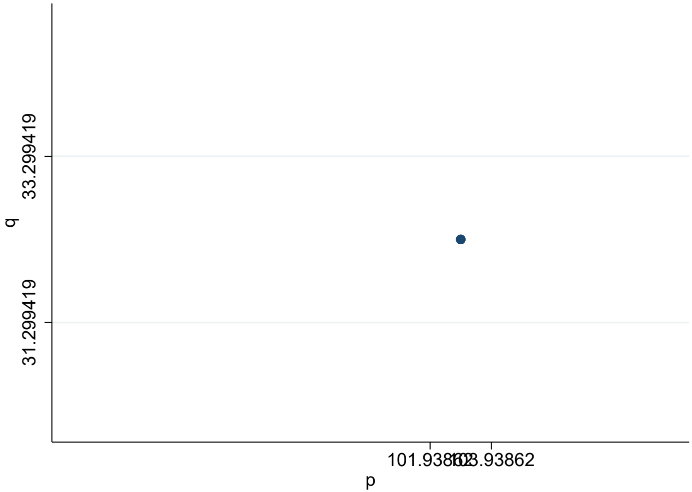
  $\rightarrow$ you cannot run experiments.

- You would still want to know demand (to calculate e.g. price cost margins, consumer surplus).

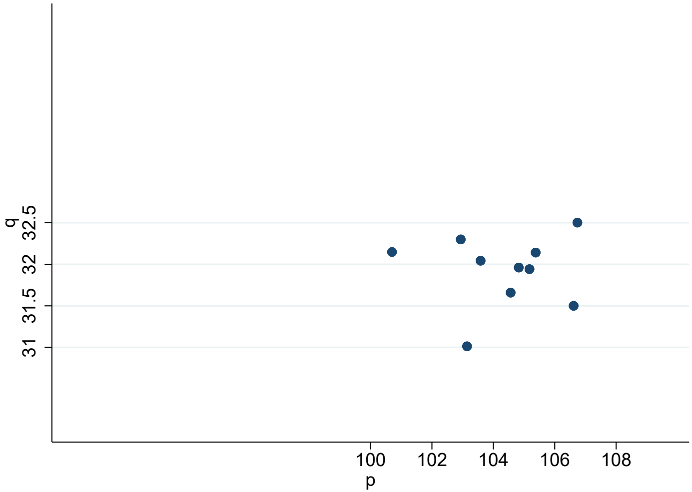
Market outcomes

- We collect data from the market.
- We observe pairs (P_i, Q_i), $i =$ market.
- Let's think how such pairs are determined, using a simple monopoly model.
- Let's allow a monopolist to choose prices instead in the same market.

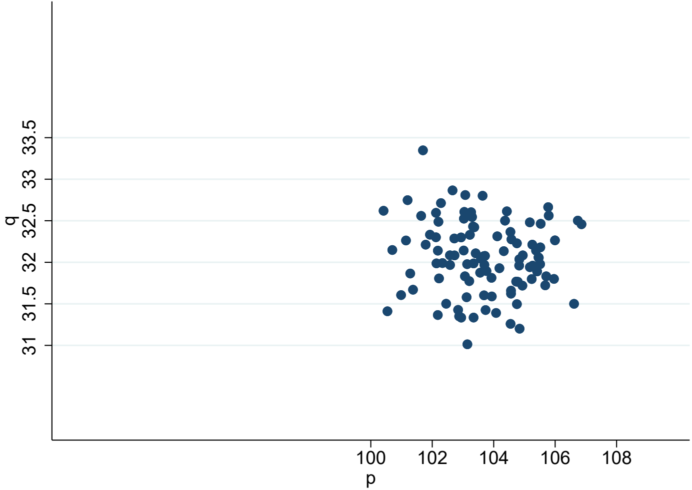
Market outcome data



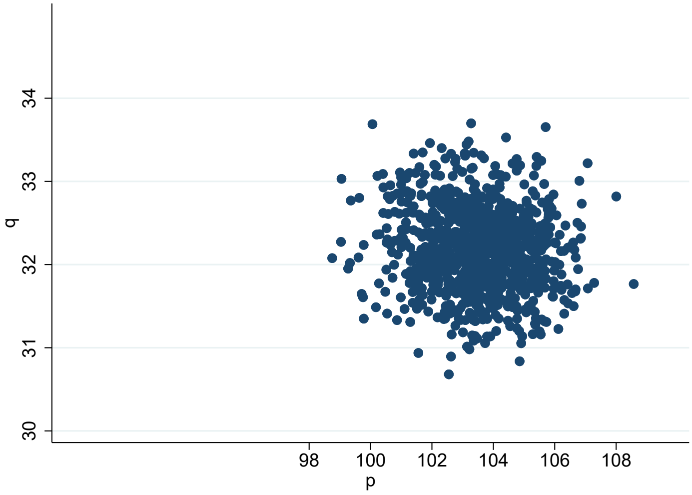
Market outcome data



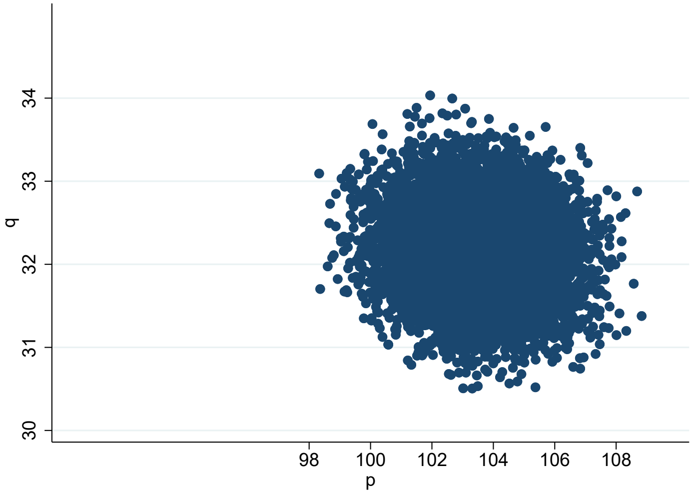
Market outcome data



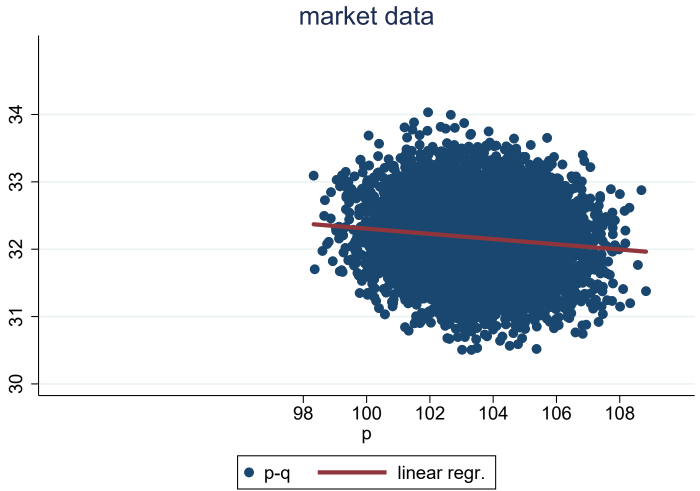
Market outcome data



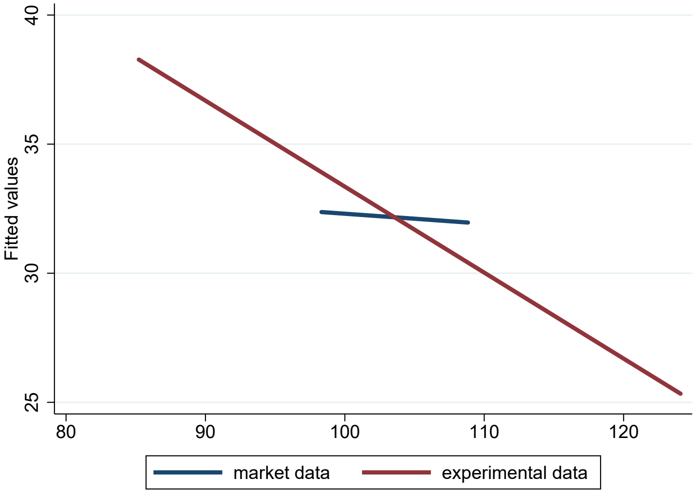
Market outcome data



Market outcome data



Market outcome data



Regression using market data

```
. regr q p

      Source |       SS           df       MS      Number of obs   =    10,000
-------------+----------------------------------   F(1, 9998)      =    135.06
       Model |  33.3509286         1  33.3509286   Prob > F        =    0.0000
    Residual |  2468.76463     9,998  .246925848   R-squared       =    0.0133
-------------+----------------------------------   Adj R-squared   =    0.0132
       Total |  2502.11556     9,999  .250236579   Root MSE        =    .49692

------------------------------------------------------------------------------
           q |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           p |  -.0386712   .0033275   -11.62   0.000    -.0451938   -.0321487
       _cons |   36.17252   .3444748   105.01   0.000     35.49728    36.84776
------------------------------------------------------------------------------
```

Parameter comparison for demand and inverse demand functions

```
. scalar alpha_ols                = -_b[_cons] / _b[p]

. scalar beta_ols                 = -1 / _b[p]

. scalar a_ols            = _b[_cons]

. scalar b_ols            = -_b[p]

. scalar list a_exp b_exp a_ols b_ols
     a_exp =  66.653405
     b_exp =  .33299579
     a_ols =  36.172515
     b_ols =  .03867121

. scalar list alpha_exp beta_exp alpha_ols beta_ols
 alpha_exp =  200.16291
  beta_exp =   3.003041
 alpha_ols =  935.38611
  beta_ols =  25.859029
```

Challenge with market data

- Price quantity pairs are a leading example of simultaneous causality.
- This generalizes to more complicated markets with:
  1. differentiated goods
  2. multiproduct firms
  3. endogenous entry and exit
  4. dynamic considerations (e.g. collusion, durable goods, ...)
  5. advertising
  6. ...

Challenge with market data

- Need to address simultaneous causality.
- $\rightarrow$ need to understand and exploit determinants of price and quantity.
- How did the experiment solve the problem?
- By having the researcher shift (=change) prices instead of the firm.

Linear monopoly

- Demand function

$$Q_i = a - bP_i + \epsilon_i$$

- $a =$ average intercept.
- $b =$ slope.
- $\epsilon_i = =$ market specific deviation from the average intercept.
- NOTE: we assume the firm observes **all** these parameters.
- Question: what if the firm did not observe our "unobservable", i.e., $\epsilon_i =?$

Linear monopoly

- Inverse demand function

$$P_i = \frac{a}{b} - \frac{1}{b}Q_i + \frac{1}{b}\epsilon_i = \alpha + \beta Q_i + \tilde{\epsilon}_i$$

How to get the supply function?

- Inverse demand function.
- Need to specify costs of production: constant marginal cost

$$c_i = c_0 + c_1 z_i + \eta_i$$

- Note: here one should have an understanding of the production technology.

How to get the supply function?

$$c_i = c_0 + c_1 z_i + \eta_i$$

- c_0 = average of marginal cost when $z_i = 0$.
- z_i = a component of marginal cost that varies across markets (cost of raw materials / unit of output, cost of labor / unit of output, ..)
- η_i = "shock" to average marginal cost / deviation from the avg. This is observed by the firm but not by the econometrician.
- If the firm did not observe η_i, how could it take it into account in its decision?

How to get the supply function?

The monopolist's problem:

$$max_{P_i} \pi_i = (P_i - c_i) \times Q_i$$

Equilibrium price:

$$P_i = \frac{a}{2b} + \frac{1}{2} c_i + \frac{1}{2b} \epsilon_i$$

$$= \frac{a}{2b} + \frac{1}{2}(c_0 + c_1 z_i + \eta_i) + \frac{1}{2b} \epsilon_i$$

How to get the supply function?

Equilibrium quantity

$$Q_i = \frac{a}{2} - \frac{b}{2}(c_0 + c_1 z_i + \eta_i) + \frac{1}{2}\epsilon_i$$

How to get the supply function?

Equilibrium price and equilibrium quantity are functions of:

How to get the supply function?

Equilibrium price and equilibrium quantity are functions of:

1. (fixed) demand parameters a and b

How to get the supply function?

Equilibrium price and equilibrium quantity are functions of:

1. (fixed) demand parameters a and b
2. (fixed) supply side parameters c_0 and c_1

How to get the supply function?

Equilibrium price and equilibrium quantity are functions of:

1. (fixed) demand parameters a and b
2. (fixed) supply side parameters c_0 and c_1
3. variable cost determinant z_i

How to get the supply function?

Equilibrium price and equilibrium quantity are functions of:

1. (fixed) demand parameters a and b
2. (fixed) supply side parameters c_0 and c_1
3. variable cost determinant z_i
4. cost shock η_i

How to get the supply function?

Equilibrium price and equilibrium quantity are functions of:

1. (fixed) demand parameters a and b
2. (fixed) supply side parameters c_0 and c_1
3. variable cost determinant z_i
4. cost shock η_i
5. demand shock ϵ_i

Market data challenge

- Both eq. price and eq. quantity are functions of:

1. demand shock ϵ_i
2. supply (cost) shock η_i

$\rightarrow$ simultaneous causality.

$\rightarrow$ price is an endogenous variable (ϵ_i is the "omitted" variable that affects both price and quantity).

Market data solution

- We want to learn the demand curve.
- Could we mimic the experimental approach with market data?
- Needed: something that shifts firm's (supply) decision at "random".
- Random = without being affected by demand shock ϵ_i.

Experiment #2

- Imagine the firm still sets the price,
- But we choose (=randomize) $z_i = z_i^{exp}$.
- Recall that $c_i = c_0 + c_1 z_i^{exp} + \eta_i$.

  $\rightarrow$ we "shift" firm's marginal cost.

Experiment #2

- Now the firm sets each time the price

$$P_i = \frac{a}{2b} + \frac{1}{2}(c_0 + c_1 z_i^{exp} + \eta_i) + \frac{1}{2b}\epsilon_i$$

$\rightarrow$ equivalent to running an experiment.

- Change in price due to (known) change in z_i^{exp}.

Experiment #2

- Imagine we raise z_i^{exp} by 1 unit.
- By how much does
  1. marginal cost $c_i = c_0 + c_1 z_i + \eta_i$ change? Answer: c_1.
  2. price change? Answer: $\frac{1}{2}c_1$ (by the equilibrium price equation).
  3. demand change? Answer: $-b\frac{1}{2}c_1$ (by the equilibrium quantity equation).

Experiment #2

- How could we get the slope of the demand function from these changes?
- Yes, by dividing the change in quantity by the change in demand:

$$-\frac{b\frac{1}{2}c_1}{\frac{1}{2}c_1} = -b$$

Experiment #2

- How could we get those numbers?

1. Regress P_i on z_i^{exp} to get $\frac{1}{2}c_1$.

$$P_i = \gamma_0 + \gamma_1 z_i + e_i$$

$$\gamma_1 = \frac{1}{2}c_i$$

Experiment #2

2. Regress Q_i on z_i^{exp} to get $-b\frac{1}{2}c_1$.

$$Q_i = \mu_0 + \mu_1 z_i + w_i$$

$$\mu_1 = -b\frac{1}{2}c_i$$

- The are called reduced form equations.

Experiment #2

- The regression equations we estimated, i.e.,

$$P_i = \gamma_0 + \gamma_1 z_i + e_i$$

$$Q_i = \mu_0 + \mu_1 z_i + w_i$$

- are called **reduced form** equations.

What are reduced form equations?

- Proper definition: a **reduced form equation** is an equation whose parameters are functions of the **structural** parameters.
- In our model, structural parameters are a, b, c_0 and c_1.
  1. They are building blocks of the theory model
  2. They are determined outside our model
  3. They are not functions of any other parameters (or variables) of the model
- The parameters (γ_0, γ_1, μ_0, μ_1) of the two regressions (P_i on z_i and Q_i on z_i) we ran are functions of the structural parameters.

What are reduced form equations?

- Commonly used meaning: a **reduced form equation** is an equation that is not derived from a theoretical model.
- Example: The regressions in the papers we have studied, i.e.,
  - Bronnenberg et al., 2015.
  - Kleven et al., 2011.

Experiment #2

- When would this work?

  1. z_i^{exp} **has to have** an impact on the decision of the firm, i.e., have an effect on c_i.

     $\rightarrow c_1$ cannot be (insignificantly different from) zero.

  2. z_i^{exp} **may not have** an effect on Q_i directly, but only via c_i.

Let's regress P on z.

```
. regr p z

      Source |       SS           df       MS      Number of obs   =    10,000
-------------+----------------------------------   F(1, 9998)      =   7814.03
       Model |  9783.4964         1  9783.4964     Prob > F        =    0.0000
    Residual |  12517.9151     9,998  1.25204192   R-squared       =    0.4387
-------------+----------------------------------   Adj R-squared   =    0.4386
       Total |  22301.4115     9,999  2.23036418   Root MSE        =    1.1189

------------------------------------------------------------------------------
           p |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           z |   .4986242   .0056407    88.40   0.000     .4875673    .5096812
       _cons |   101.0138   .0304066  3322.11   0.000     100.9542    101.0734
------------------------------------------------------------------------------

. scalar red_1    = _b[z]
```

Let's regress Q on z.

```
. regr q z

      Source |       SS           df       MS      Number of obs   =    10,000
-------------+----------------------------------   F(1, 9998)      =   8117.89
       Model |  1121.21997         1  1121.21997   Prob > F        =    0.0000
    Residual |  1380.89558     9,998  .138117182   R-squared       =    0.4481
-------------+----------------------------------   Adj R-squared   =    0.4481
       Total |  2502.11556     9,999  .250236579   Root MSE        =    .37164

------------------------------------------------------------------------------
           q |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           z |  -.1687997   .0018735   -90.10   0.000    -.1724721   -.1651273
       _cons |   33.01561   .0100991  3269.17   0.000     32.99581     33.0354
------------------------------------------------------------------------------

. scalar red_2    = _b[z]
```

Let's calculate b.

```
. scalar b_red    = red_2 / red_1

. scalar list b_red
    b_red = -.33853094
```

Instrumental variable

- **Instrumental variable** = a variable that **causes variation** in price (explanatory variable X) but **does not affect** demand (dependent variable Y) directly.
- If the variable cost component z_i varies "at random", i.e., without affecting demand directly,

  $\rightarrow$ market data allows us to use the “experimental approach” indirectly.

Approach #2

- Could we proceed differently?
  1. Regress P_i on z_i. Calculate predicted price $\hat{P}_i = \hat{\gamma}_0 + \hat{\gamma}_1 z_i$.
  2. Regress Q_i on $\hat{P}_i$ to get b (and a).
- Equation $Q_i = a - bP_i + \epsilon_i$ is a **structural** equation. Why?

Approach #2

- Could we proceed differently?

  1. Regress P_i on z_i. Calculate predicted price $\hat{P}_i = \hat{\gamma}_0 + \hat{\gamma}_1 z_i$.
  2. Regress Q_i on $\hat{P}_i$ to get b (and a).

- Equation $Q_i = a - bP_i + \epsilon_i$ is a **structural** equation. Why?
- Because it is a function of structural parameters only ($+$ P_i which is determined within the model).

Approach #2

- The parameters of a structural equation are part of the **model primitives**, i.e.,
  - they are not determined within the model
  - they are not functions of other parameters of the model
- **Reduced form parameters** are functions of structural parameters.

Regress P on z, create predicted values

```
. regr p z

      Source |       SS           df       MS      Number of obs   =    10,000
-------------+----------------------------------   F(1, 9998)      =   7814.03
       Model |  9783.4964         1   9783.4964   Prob > F        =    0.0000
    Residual |  12517.9151     9,998  1.25204192   R-squared       =    0.4387
-------------+----------------------------------   Adj R-squared   =    0.4386
       Total |  22301.4115     9,999  2.23036418   Root MSE        =    1.1189

------------------------------------------------------------------------------
           p |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           z |   .4986242   .0056407    88.40   0.000     .4875673    .5096812
       _cons |   101.0138   .0304066  3322.11   0.000     100.9542    101.0734
------------------------------------------------------------------------------

. predict p_hat
(option xb assumed; fitted values)
```

Regress Q on $\hat{P}$

```
. regr q p_hat

      Source |       SS           df       MS      Number of obs   =    10,000
-------------+----------------------------------   F(1, 9998)      =   8117.89
       Model |  1121.21997         1  1121.21997   Prob > F        =    0.0000
    Residual |  1380.89558     9,998  .138117182   R-squared       =    0.4481
-------------+----------------------------------   Adj R-squared   =    0.4481
       Total |  2502.11556     9,999  .250236579   Root MSE        =    .37164

------------------------------------------------------------------------------
           q |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       p_hat |  -.3385309   .0037573   -90.10   0.000     -.345896   -.3311659
       _cons |   67.21192   .3889483   172.80   0.000      66.4495    67.97433
------------------------------------------------------------------------------
```

Why / how do these approaches work?

1. Regress P_i on z_i to get $\frac{1}{2}c_1 = \frac{cov(P_i,z_i)}{var(z_i)}$.
2. Regress Q_i on z_i to get $-b\frac{1}{2}c_1 = \frac{cov(Q_i,z_i)}{var(z_i)}$.

$\rightarrow -b = \frac{cov(Q_i,z_i)}{cov(P_i,z_i)}$.

Regress Q on $\hat{P}$

$$b = \frac{cov(Q_i, \hat{P}_i)}{var(\hat{P}_i)} = \frac{cov(Q_i, \hat{\gamma}_0 + \hat{\gamma}_1 z_i)}{var(\hat{\gamma}_0 + \hat{\gamma}_1 z_i)}$$

$$\rightarrow b = \frac{\hat{\gamma}_1 cov(Q_i, z_i)}{\hat{\gamma}_1^2 var(z_i)} = \frac{1}{\hat{\gamma}_1} \frac{cov(Q_i, z_i)}{var(z_i)}$$

$$\text{because } \hat{\gamma}_1 = \frac{cov(P_i, z_i)}{var(z_i)} \rightarrow$$

$$b = \frac{var(z_i)}{cov(P_i, z_i)} \frac{cov(Q_i, z_i)}{var(z_i)} = \frac{cov(Q_i, z_i)}{cov(P_i, z_i)}$$

2SLS / instrumental variables regression

- In practice, want to use the so called Two Stage Least Squares (2SLS) or instrumental variables regression command. In Stata, `ivregress` or from SSC `ivreg2`.
- Manual and `ivregress` command(s) produce same point estimates, but the latter corrects the standard errors.
- This is important, as the manual approach yields too small standard errors: It ignores the uncertainty in the parameters $\hat{\gamma}_0$ and $\hat{\gamma}_1$ used to calculate $\hat{P}_i$.

2SLS estimation of demand

```
. ivregress 2sls q (p = z)

Instrumental variables (2SLS) regression          Number of obs   =     10,000
                                                  Wald chi2(1)    =    2506.07
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =          .
                                                  Root MSE        =     .66888

------------------------------------------------------------------------------
           q |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           p |  -.3385309   .0067624   -50.06   0.000     -.351785   -.3252769
       _cons |   67.21192   .7000301    96.01   0.000     65.83988    68.58395
------------------------------------------------------------------------------
Instrumented:  p
Instruments:   z
```

Requirements for an instrument

- Think of our normal regression $Y = \beta_0 + \beta_1 X + u$.

1. Instrument relevance: The instrument Z has to affect the (endogenous) explanatory variable X of the equation of interest ("2nd stage equation") in the equation

   $X = \alpha_0 + \alpha_1 Z + v$.

2. Instrument exogeneity: The instrument Z may not be correlated with the error term of the equation of interest, i.e.,

   $cov(Z, u) = 0$.

Instrument relevance / Weak instrument

- Relevance = instrument Z needs to be "correlated enough" with the endogenous explanatory variable X.
- What happens when $cov(Z, X) \to 0$?
- β_1 becomes undefined!

Instrument relevance / Weak instrument

$\rightarrow$ you want to check that your instrument is relevant.

$=$ you don't have a **weak instrument**.

- Rule of thumb: F-statistic of Z when you regress X on Z (and possible further controls) > 10.
- Note: with 1 instrument, F-test is the square of the t-test.
- Notice that the test for weak instruments is stricter than our usual 5% confidence level (t-stat 2).

2SLS estimation

```
. estat firststage

  First-stage regression summary statistics
  --------------------------------------------------------------------------
               |            Adjusted      Partial
      Variable |   R-sq.       R-sq.        R-sq.     F(1,9998)   Prob > F
  -------------+------------------------------------------------------------
             p |  0.4387      0.4386       0.4387       7814.03    0.0000
  --------------------------------------------------------------------------


  Minimum eigenvalue statistic = 7814.03

  Critical Values                      # of endogenous regressors:    1
  Ho: Instruments are weak             # of excluded instruments:     1
  ---------------------------------------------------------------------
                                     |    5%     10%     20%     30%
  2SLS relative bias                 |         (not available)
  -----------------------------------+---------------------------------
                                     |   10%     15%     20%     25%
  2SLS Size of nominal 5% Wald test  |  16.38    8.96    6.66    5.53
  LIML Size of nominal 5% Wald test  |  16.38    8.96    6.66    5.53
  ---------------------------------------------------------------------
```

Instrument relevance / Weak instrument

- There are more sophisticated tests.
- There are ways of allowing for weak instruments.
- We leave all that for later courses.
- Good instruments are hard to find...

Instrument correlated with error

$=$ **"exogeneity"** assumption or **exclusion restriction**:

$$\mathbb{E}[u|\boldsymbol{X}] = 0$$

If this condition does not hold $\rightarrow$ biased estimate of β_1.

- Similar to omitted variable bias.

Instrument correlated with error

- What can be done?
    1. Strong story for why no correlation between instrument and error.
    2. With multiple instruments, may do tests.
    3. There are ways of allowing for (some) correlation to check robustness of your results (for later).