MIT 14.32/14.320 **Spring 2023**
J. Angrist

Lecture Note 15: Simultaneous Equations Models

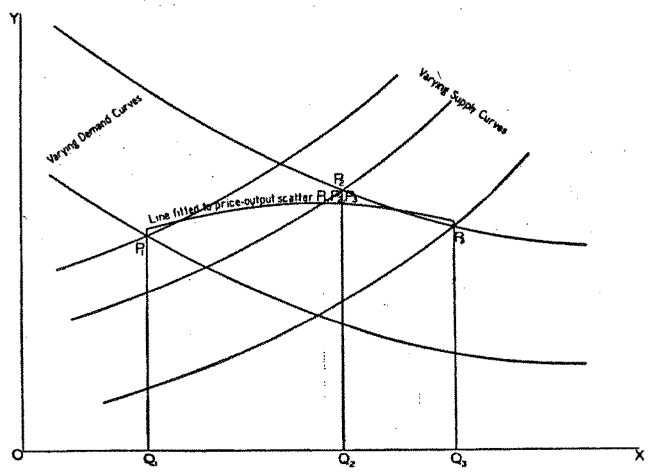
1 All Together Now

- Some economic policy questions, like effects of changing tax rates, hinge on elasticities of demand and supply (can you think of other policies besides taxes for which elasticities matter?). How should supply and demand elasticities should be estimated?

- A regression of quantity (q_t) on price (p_t) across markets indexed by t gives the best (MMSE) approximation to $E[q_t \mid p_t]$ for *equilibrium* quantities and prices, i.e., those that equate supply and demand. The figure below suggests that the regression of quantity on price characterizes neither the supply nor demand function:

If both supply and demand conditions change, price-output data yield no direct information as to either curve. (Figure 4.)

FIGURE 4. PRICE-OUTPUT DATA FAIL TO REVEAL EITHER SUPPLY OR DEMAND CURVE.



- This figure comes from the appendix to a 1928 study by once-obscure economist (and aspiring poet) Phillip Wright called *The Tariff on Animal and Vegetable Oils*. Little-noticed at the time, Wright's "Appendix B" laid the intellectual foundations of modern econometrics.

- Wright's problem is not *statistical*. Rather, just as in the story of short and long regressions, and in regression models with mismeasured regressors, the regression we've got is not the one we want. Here, however, the "regression we want" is not a regression at all, but rather a theoretical economic relationship that describes the counterfactual choices of buyers and sellers.
  - Economists refer to theoretical causal relationships of the sort described by supply and demand curves as *structural equations*.

- Applied econometricians interested in uncovering structural relationships are said to face an *identification problem.* The term “identification,” which we’ve already encountered in the context of IV, originates with study of simultaneous equations models (SEMs).

1.1 Simultaneous Equations Bias

- When prices and quantities are determined as the solution to a simultaneous equations system, OLS estimates are not consistent for supply and demand elasticities; this bad behavior is called *simultaneous equations bias.*
  - Not an obvious conclusion: generations of statisticians have been confused by this[1]
- Economists use simplifying assumptions to understand and solve econometric problems. The linear simultaneous equations model (SEM) is one such simplification. This model asserts that potential quantities supplied and demanded can be written as a linear function of prices and possibly other variables that we think of as shifting these functions. Specifically, we write the structural equations of interest as:

$$q_t^d(p) = \alpha_0 + \alpha_1 p + \alpha_2 z_t + \epsilon_t^d \tag{1}$$
$$q_t^s(p) = \beta_0 + \beta_1 p + \beta_2 x_t + \epsilon_t^s, \tag{2}$$

  where $q_t^d(p)$ is the quantity consumers demand at price p, $q_t^s(p)$ is the quantity producers supply at price p, and z_t and x_t are additional observed determinants of demand and supply (such as consumer income and prices of other goods). The market equilibrium price, p_t, solves:

$$q_t^d(p_t) = q_t^s(p_t) = q_t \tag{3}$$

- Variables determined by solving a system of equations are said to be *endogenous*, meaning their values are determined within the system. Variables like z_t and x_t, determined outside the system, are said to be *exogenous.*
- What do OLS estimates of (1) or (2) produce when the observed p_t and q_t satisfy equations (1)-(3)?
  - To answer this, we first solve for the *reduced form* for p_t by equating supply and demand:

$$\begin{aligned} p_t &= \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} - \frac{\alpha_2}{\alpha_1 - \beta_1} z_t + \frac{\beta_2}{\alpha_1 - \beta_1} x_t + \frac{\epsilon_t^s - \epsilon_t^d}{\alpha_1 - \beta_1} \\ &= \pi_{10} + \pi_{11} z_t + \pi_{12} x_t + \nu_{1t} \end{aligned} \tag{4}$$

    Note that the random part of p_t – error term ν_{1t} – is surely correlated with the structural errors in (1) and (2)
- From this, we learn that structural equations in a simultaneous equations model are <u>not</u> regressions, so OLS does not reliably estimate the coefficients in them.
- Old-school SEMs are not much seen in modern empirical work, but Wright’s analysis of the SEM is of incomparable intellectual importance. The SEM is the foundation upon which our modern ’metrics house is built. Encountering Wright’s elegant framework as a college sophomore in 1980, I was floored. Today, I appreciate it even more, since Wright gave birth to IV.

[1] See, e.g., Dawid (1984) and Cox (1992) cited in Imbens’ (2022) Nobel Lecture.

2 Identifying Structural Equations from Reduced Forms

- The reduced form for p_t is (4). The reduced form for q_t is shown below (derive this):

$$\begin{aligned} q_t &= \frac{\beta_1\alpha_0 - \alpha_1\beta_0}{\beta_1 - \alpha_1} + \frac{\alpha_2\beta_1}{\beta_1 - \alpha_1} z_t - \frac{\alpha_1\beta_2}{\beta_1 - \alpha_1} x_t + \frac{\beta_1\epsilon_t^s - \alpha_1\epsilon_t^d}{\beta_1 - \alpha_1} \\ &= \pi_{20} + \pi_{21} z_t + \pi_{22} x_t + \nu_{2t} \end{aligned} \tag{5}$$

- In contrast with structural equations, reduced form equations are regressions: their errors are uncorrelated with all RHS variables. This is because regressors in the reduced form (x_t and z_t) are assumed to be uncorrelated with the structural errors. These *exogenous variables* are determined outside the system. In general, reduced form equations (one for each endogenous variable) are found by solving for each endogenous variable as a function of all of the exogenous variables in the system.

- *When is an SEM identified?* When the structural coefficients can be obtained from the reduced form coefficients.
  - We can manipulate the reduced form coefficients in (4) and (5) to obtain the structural coefficients of interest; try this and see
  - But ponder this riff on (1) and (2):

$$q_t^d(p) = \alpha_0 + \alpha_1 p + \alpha_2 z_t + \alpha_3 x_t + \epsilon_t^d \tag{6}$$
$$q_t^s(p) = \beta_0 + \beta_1 p + \beta_2 z_t + \beta_3 x_t + \epsilon_t^s \tag{7}$$

  This system is *under-identified*: knowledge of the corresponding reduced-form parameters does not reveal the underlying structure in this case (show this)

- Identification in the SEM requires *exclusion restrictions*; verily, identification requires instruments!

- **Heads-up on the lingo**:
  - The term *reduced form*, which we've long been using, originates in early analyses of the SEM
  - In the SEM, there's a reduced form for each endogenous variable, including the endogenous variables found on the right-hand side of the structural equations. In the SEM, therefore, the *reduced form* for price is also the *first stage* for price when we compute 2SLS estimates of demand and supply elasticities.
    * When doing 2SLS outside an SEM framework, it's useful to distinguish first-stage equations from other reduced form equations.

Structural vs Reduced-Form Policy Analysis

- Structural equations are models for potential outcomes: they tell us what would happen under hypothetical random assignment of right-hand-side variables. Read all about it in The Fish Paper.

- As we've seen, some policy questions can be tackled head-on: Card and Krueger (1994) estimate minimum wage effects without benefit of a structural model. Theirs is sometimes said to be a "reduced form policy analysis" because it looks at minimum wage effects directly, rather than through an SEM characterizing equilibrium in the labor market.

- Supply and demand elasticities are often of intrinsic interest, as in the Graddy (1995) study of price discrimination. Do Asians pay less for fish because their demand for fish is more elastic? We need structural elasticities of demand (for fish) to see.

- Many policy questions require extrapolation beyond the data at hand. This was once the case with a $15 minimum wage, a level seen in the US only recently. Structural parameters like labor demand elasticities can be used to predict the consequences of imagined but as-yet-unseen changes.

3 Estimating Simultaneous Equations Models

- OLS estimates of structural equations are neither unbiased nor consistent for structural parameters.
- What is? Indirect least squares (ILS), instrumental variables (IV), and two-stage least squares (2SLS) estimates, to name three (not a complete list).

3.1 Indirect Least Squares

- OLS estimates of reduced form coefficients (the π's, above) are consistent because reduced forms are regressions.
- Indirect least squares (ILS) estimators solve for structural coefficients from reduced form estimates ... *if the model is identified.* Using (4) and (5), we have:

$$\pi_{11} = \frac{-\alpha_2}{\beta_1 - \alpha_1} \qquad \pi_{21} = \frac{-\beta_1\alpha_2}{\beta_1 - \alpha_1}$$
$$\pi_{12} = \frac{\beta_2}{\beta_1 - \alpha_1} \qquad \pi_{22} = \frac{\beta_2\alpha_1}{\beta_1 - \alpha_1},$$

so that

$$\frac{\pi_{21}}{\pi_{11}} = \beta_1 \qquad \frac{\pi_{22}}{\pi_{12}} = \alpha_1 \tag{8}$$

We can also solve for the structural coefficients on exogenous variables as:

$$-\pi_{12}(\beta_1 - \alpha_1) = \beta_2 \qquad -\pi_{11}(\beta_1 - \alpha_1) = \alpha_2$$

- ILS estimators substitute sample analogs for population π's in the above formulas
  - ILS estimates of structural parameters are consistent (why?)
  - ILS is IV (see if you can figure out what the instruments are by simplifying the ratio of reduced form coefficients in 8)

3.2 Two-Stage Least Squares (2SLS) Reprise

- 2SLS logic: substitute first stage fitted values for RHS endogenous variables
  - Consider a simple system with one exogenous variable:

$$q_t^d = \alpha_0 + \alpha_1 p_t + \epsilon_t^d \tag{9}$$
$$q_t^s = \beta_0 + \beta_1 p_t + \beta_2 x_t + \epsilon_t^s \tag{10}$$

  The reduced form (first stage) for price is:

$$p_t = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{\beta_2}{\alpha_1 - \beta_1}x_t + \frac{\epsilon_t^s - \epsilon_t^d}{\alpha_1 - \beta_1} \tag{11}$$
$$= \pi_{10} + \pi_{11}x_t + \nu_{1t}$$

- Use this and equation (9) to write quantity demanded as:

$$q_t = \alpha_0 + \alpha_1 \hat{p}_t + [\epsilon_t^d + (p_t - \hat{p}_t)\alpha_1], \tag{12}$$

where

$$\hat{p}_t = \hat{\pi}_{10} + \hat{\pi}_{11} x_t$$

- Note that $\hat{p}_t$ is a linear function of x_t, which is uncorrelated with ϵ_t^d. Also, $\hat{p}_t$ is necessarily uncorrelated with $(p_t - \hat{p}_t)$. Thus, OLS estimates of α_1 in (12) are consistent
  - OLS regression on the fitted values generated by a first stage regression is the *2SLS second stage*
- In a model with one instrument and one endogenous variables, as in (12), 2SLS estimates are identical to the corresponding ILS and IV estimates (LN11 shows this)

3.3 Over-identified Models

- Up your game by adding an extra instrument (exogenous variable) to the supply equation:

$$q_t^d(p) = \alpha_0 + \alpha_1 p_t + \epsilon_t^d \tag{13}$$

$$q_t^s(p) = \beta_0 + \beta_1 p_t + \beta_2 x_t + \beta_3 z_t + \epsilon_t^s \tag{14}$$

The demand equation is now *over-identified,* meaning we have more than one ILS solution for the demand slope, α_1, and more than one instrument available for IV and 2SLS estimation of α_1

  - To see this, note that the reduced forms for this system are:

$$\begin{aligned} p_t &= \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{\beta_2}{\alpha_1 - \beta_1} x_t + \frac{\beta_3}{\alpha_1 - \beta_1} z_t - \frac{\epsilon_t^s - \epsilon_t^d}{\alpha_1 - \beta_1} \\ &= \pi_{10} + \pi_{11} x_t + \pi_{12} z_t + \nu_{1t} \\ q_t &= \frac{\beta_1 \alpha_0 - \alpha_1 \beta_0}{\beta_1 - \alpha_1} - \frac{\alpha_1 \beta_2}{\beta_1 - \alpha_1} x_t - \frac{\alpha_1 \beta_3}{\beta_1 - \alpha_1} z_t + \frac{\beta_1 \epsilon_t^s - \alpha_1 \epsilon_t^d}{\beta_1 - \alpha_1} \\ &= \pi_{20} + \pi_{21} x_t + \pi_{22} z_t + \nu_{2t} \end{aligned}$$

  These reduced forms generate two solutions for α_1:

$$\alpha_1 = \frac{\pi_{21}}{\pi_{11}} \text{ or } \alpha_1 = \frac{\pi_{22}}{\pi_{12}}$$

- We also have two instruments available to estimate the demand slope:

$$\begin{aligned} \alpha_1 &= \frac{\pi_{21}}{\pi_{11}} = \frac{C(q_t, \tilde{x}_t)}{C(p_t, \tilde{x}_t)} \\ &= \frac{\pi_{22}}{\pi_{12}} = \frac{C(q_t, \tilde{z}_t)}{C(p_t, \tilde{z}_t)} \end{aligned}$$

- 2SLS combines the two instruments, $\tilde{x}_t$ and $\tilde{z}_t$, and the ILS/IV estimates they generate (what's the $\tilde{}$ on top do here?)
  - Use OLS to estimate the first stage fitted values in a regression with both exogenous variables on the RHS:

$$\begin{aligned} p_t &= \pi_{10} + \pi_{11} x_t + \pi_{12} z_t + \nu_{1t} \\ \hat{p}_t &= \hat{\pi}_{10} + \hat{\pi}_{11} x_t + \hat{\pi}_{12} z_t \end{aligned}$$

- The 2SLS estimate of α_1 is then the sample analog of

$$\frac{C(q_t,\,\hat{p}_t)}{V(\hat{p}_t)} = OLS \text{ slope for } q_t \text{ on } \hat{p}_t = \frac{C(q_t,\,\hat{p}_t)}{C(p_t,\,\hat{p}_t)} = IV \text{ using } \hat{p}_t \text{ as an instrument}$$

- Note that the single instrument, $\hat{p}_t$, is a linear combination of the two instruments, x_t and z_t (and so of $\tilde{x}_t$ and $\tilde{z}_t$)
  - In fact, 2SLS is the efficient IV estimator for an over-identified homoskedastic model.
  - That is, the first-stage fitted value is the best instrument among all possible instruments constructed as a linear combination of the multiple instruments at hand ("best" means the IV estimator that yields the most precise structural coefficient estimate. Pretty neat!)
- The Song Remains the Same: when *doing* the SEM, we do 2SLS! (Led Zep said it best)
  - 2SLS estimates of over-identified models are statistically efficient, but 2SLS estimates of under-identified models do not exist; test your SEM understanding by showing this second point

4 Something Fishy at the Fulton Fish Market

- Graddy (1995) and Angrist, Graddy, and Imbens (2000) compute 2SLS estimates of the elasticity of demand for fish using this model:

$$q_t^d(p_t) = \alpha_0 + \alpha_1 p_t + \alpha_2' x_t + \epsilon_t^d \tag{15}$$
$$q_t^s(p_t) = \beta_0 + \beta_1 p_t + \beta_2' x_t + \beta_3' z_t + \epsilon_t^s \tag{16}$$
$$q_t^d = q_t^s = q_t$$

- The price reduced form is the relevant first stage. This can be written:

$$p_t = \pi_{10} + \pi_{11}' x_t + \pi_{12}' z_t + \nu_{1t},$$

  where

$$\begin{aligned} q_t &= \text{daily quantity demanded from a Fulton dealer} \\ p_t &= \text{average daily price} \\ x_t &= \text{vector of day-of-the-week dummies} \\ z_t &= \text{vector of measures of weather conditions at sea (wind speed and wave height)} \end{aligned}$$

  What makes the resulting 2SLS estimates a demand elasticity rather than a supply elasticity? Exclusion!

- Annoying details
  1. Data available for only one dealer
  2. No information available on prices of substitute goods

```
. gen mixed3=(1-stormy3)*(speed3>15)*(wave3>3)

. gen stormy2=(speed2>12)*(wave2>5.5)

. gen mixed2=(1-stormy2)*(speed2>10)*(wave2>3)

.
. // OLS and reduced form estimates
. reg lnqty day1 day2 day3 day4 lnprice

      Source |       SS       df       MS              Number of obs =      97
-------------+------------------------------           F(  5,    91) =    5.04
       Model |  12.1722085     5  2.4344417            Prob > F      =  0.0004
    Residual |   43.960225    91  .483079395           R-squared     =  0.2168
-------------+------------------------------           Adj R-squared =  0.1738
       Total |  56.1324335    96  .584712849           Root MSE      =  .69504

------------------------------------------------------------------------------
       lnqty |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        day1 |  -.3109272   .2258233    -1.38   0.172    -.7594975    .1376431
        day2 |  -.6827901    .222667    -3.07   0.003    -1.125091   -.2404896
        day3 |  -.5338939   .2199374    -2.43   0.017    -.9707725   -.0970152
        day4 |   .0672273   .2204205     0.30   0.761     -.370611    .5050656
     lnprice |  -.5246553   .1761115    -2.98   0.004    -.8744792   -.1748314
       _cons |   8.244317   .1628134    50.64   0.000     7.920909    8.567726
------------------------------------------------------------------------------

. reg lnqty day1 day2 day3 day4 speed3

      Source |       SS       df       MS              Number of obs =      97
-------------+------------------------------           F(  5,    91) =    4.77
       Model |  11.6510004     5  2.33020008           Prob > F      =  0.0006
    Residual |  44.4814331    91  .488806957           R-squared     =  0.2076
-------------+------------------------------           Adj R-squared =  0.1640
       Total |  56.1324335    96  .584712849           Root MSE      =  .69915

------------------------------------------------------------------------------
       lnqty |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        day1 |  -.2669628   .2278686    -1.17   0.244    -.7195958    .1856702
        day2 |  -.5323006   .2301553    -2.31   0.023    -.9894759   -.0751253
        day3 |  -.4803295   .2228455    -2.16   0.034    -.9229846   -.0376743
        day4 |   .0049691   .2211368     0.02   0.982     -.434292    .4442301
      speed3 |  -.0316299   .0113951    -2.78   0.007    -.0542647    -.008995
       _cons |   8.999321    .269828    33.35   0.000     8.463341    9.535302
------------------------------------------------------------------------------

. reg lnprice day1 day2 day3 day4 speed3

      Source |       SS       df       MS              Number of obs =      97
-------------+------------------------------           F(  5,    91) =    2.17
       Model |   1.6717106     5   .33434212           Prob > F      =  0.0646
    Residual |  14.0413996    91  .154301095           R-squared     =  0.1064
-------------+------------------------------           Adj R-squared =  0.0573
```

```
       Total |  15.7131102    96  .163678231           Root MSE      =  .39281

------------------------------------------------------------------------------
     lnprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        day1 |   -.020083   .1280266    -0.16   0.876    -.2743922    .2342262
        day2 |  -.1004775   .1293114    -0.78   0.439    -.3573387    .1563837
        day3 |  -.0038505   .1252044    -0.03   0.976    -.2525537    .2448527
        day4 |   .1026251   .1242444     0.83   0.411    -.1441711    .3494214
      speed3 |   .0201874   .0064022     3.15   0.002     .0074701    .0329046
       _cons |  -.6651257   .1516013    -4.39   0.000     -.966263   -.3639884
------------------------------------------------------------------------------

. predict phat1, xb

.
. reg lnqty day1 day2 day3 day4 mixed3 stormy3

      Source |       SS       df       MS              Number of obs =      97
-------------+------------------------------           F(  6,    90) =    4.31
       Model |  12.5323209     6  2.08872015           Prob > F      =  0.0007
    Residual |  43.6001126    90  .484445695           R-squared     =  0.2233
-------------+------------------------------           Adj R-squared =  0.1715
       Total |  56.1324335    96  .584712849           Root MSE      =  .69602

------------------------------------------------------------------------------
       lnqty |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        day1 |  -.3171298    .229444    -1.38   0.170    -.7729603    .1387007
        day2 |  -.6655481   .2230281    -2.98   0.004    -1.108632   -.2224639
        day3 |  -.5577861   .2206741    -2.53   0.013    -.9961936   -.1193785
        day4 |  -.0345707    .221351    -0.16   0.876    -.4743231    .4051817
      mixed3 |  -.2615067   .1778435    -1.47   0.145    -.6148239    .0918104
     stormy3 |  -.5222825   .1686905    -3.10   0.003    -.8574155   -.1871494
       _cons |   8.650079   .1794426    48.21   0.000     8.293585    9.006573
------------------------------------------------------------------------------

. reg lnprice day1 day2 day3 day4 mixed3 stormy3

      Source |       SS       df       MS              Number of obs =      97
-------------+------------------------------           F(  6,    90) =    3.36
       Model |  2.87271871     6  .478786451           Prob > F      =  0.0050
    Residual |  12.8403915    90  .142671017           R-squared     =  0.1828
-------------+------------------------------           Adj R-squared =  0.1283
       Total |  15.7131102    96  .163678231           Root MSE      =  .37772

------------------------------------------------------------------------------
     lnprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        day1 |  -.0094215   .1245151    -0.08   0.940    -.2567924    .2379495
        day2 |  -.0177709   .1210333    -0.15   0.884    -.2582247    .2226828
        day3 |   .0368483   .1197558     0.31   0.759    -.2010675    .2747642
        day4 |   .1245099   .1201232     1.04   0.303    -.1141358    .3631555
      mixed3 |   .2812021   .0965125     2.91   0.005     .0894632    .4729409
     stormy3 |   .3871992   .0915453     4.23   0.000     .2053286    .5690699
       _cons |  -.4866297   .0973803    -5.00   0.000    -.6800926   -.2931668
```

```
------------------------------------------------------------------------------

. predict phat2, xb

.
. // 2nd-stage estimates -- speed as instrument
. reg lnqty phat1 day1 day2 day3 day4

      Source |       SS       df       MS              Number of obs =      97
-------------+------------------------------           F(  5,    91) =    4.77
       Model |  11.6510004     5  2.33020007           Prob > F      =  0.0006
    Residual |  44.4814331    91  .488806957           R-squared     =  0.2076
-------------+------------------------------           Adj R-squared =  0.1640
       Total |  56.1324335    96  .584712849           Root MSE      =  .69915

------------------------------------------------------------------------------
       lnqty |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       phat1 |  -1.566815   .5644648    -2.78   0.007    -2.688055    -.445575
        day1 |  -.2984291    .227249    -1.31   0.192    -.7498314    .1529731
        day2 |  -.6897303   .2240115    -3.08   0.003    -1.134702   -.2447589
        day3 |  -.4863624   .2225836    -2.19   0.031    -.9284975   -.0442274
        day4 |   .1657637   .2274403     0.73   0.468    -.2860185    .6175459
       _cons |   7.957192   .2205117    36.09   0.000     7.519173    8.395212
------------------------------------------------------------------------------

. // 2nd-stage estimates -- mixed3 and stormy3 as instruments
. reg lnqty phat2 day1 day2 day3 day4

      Source |       SS       df       MS              Number of obs =      97
-------------+------------------------------           F(  5,    91) =    5.10
       Model |  12.2836698     5  2.45673397           Prob > F      =  0.0004
    Residual |  43.8487636    91  .481854546           R-squared     =  0.2188
-------------+------------------------------           Adj R-squared =  0.1759
       Total |  56.1324335    96  .584712849           Root MSE      =  .69416

------------------------------------------------------------------------------
       lnqty |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       phat2 |  -1.268173    .419728    -3.02   0.003    -2.101911   -.4344354
        day1 |  -.3020106   .2255832    -1.34   0.184    -.7501038    .1460826
        day2 |  -.6877415    .222399    -3.09   0.003     -1.12951   -.2459732
        day3 |   -.499983    .220345    -2.27   0.026    -.9376713   -.0622948
        day4 |   .1375271   .2230704     0.62   0.539    -.3055748    .5806289
       _cons |   8.039471   .1935591    41.53   0.000      7.65499    8.423952
------------------------------------------------------------------------------

.
. // same estimates using ivreg
. version 10 // (older versions -- just use "ivreg" in place of "ivregress
2sls
> ")

. ivregress 2sls lnqty day1 day2 day3 day4 (lnprice=speed3)

Instrumental variables (2SLS) regression               Number of obs =      97
```

```
                                                      Wald chi2(5)  =   18.56
                                                      Prob > chi2   =  0.0023
                                                      R-squared     =       .
                                                      Root MSE      =  .79221

------------------------------------------------------------------------------
       lnqty |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lnprice |  -1.566815   .6395994    -2.45   0.014    -2.820407   -.3132232
        day1 |  -.2984291   .2574976    -1.16   0.246    -.8031152    .2062569
        day2 |  -.6897303   .2538292    -2.72   0.007    -1.187226   -.1922342
        day3 |  -.4863624   .2522112    -1.93   0.054    -.9806874    .0079625
        day4 |   .1657637   .2577143     0.64   0.520    -.3393472    .6708745
       _cons |   7.957192   .2498635    31.85   0.000     7.467469    8.446916
------------------------------------------------------------------------------
Instrumented:  lnprice
Instruments:   day1 day2 day3 day4 speed3

. ivregress 2sls lnqty day1 day2 day3 day4 (lnprice=mixed3 stormy3)

Instrumental variables (2SLS) regression              Number of obs =      97
                                                      Wald chi2(5)  =   22.67
                                                      Prob > chi2   =  0.0004
                                                      R-squared     =  0.0635
                                                      Root MSE      =  .73618

------------------------------------------------------------------------------
       lnqty |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lnprice |  -1.268173   .4451392    -2.85   0.004     -2.14063   -.3957166
        day1 |  -.3020106   .2392405    -1.26   0.207    -.7709133    .1668921
        day2 |  -.6877415   .2358635    -2.92   0.004    -1.150025   -.2254575
        day3 |  -.4999831   .2336852    -2.14   0.032    -.9579976   -.0419686
        day4 |   .1375271   .2365755     0.58   0.561    -.3261524    .6012066
       _cons |   8.039471   .2052776    39.16   0.000     7.637134    8.441808
------------------------------------------------------------------------------
Instrumented:  lnprice
Instruments:   day1 day2 day3 day4 mixed3 stormy3

.
. log close
       log:  /bbkinghome/paul_s/32/stataForLectures/ln18/ln18.log
  log type:  text
 closed on:  28 Apr 2009, 11:47:05
------------------------------------------------------------------------------
```

- ELASTICITY ESTIMATES BY ETHNICITY

```
. label var lnprice_w "log WHITE average price"

.
. // generate instruments
. gen stormy3=(speed3>18)*(wave3>4.5)

. gen mixed3=(1-stormy3)*(speed3>15)*(wave3>3)

. gen stormy2=(speed2>12)*(wave2>5.5)

. gen mixed2=(1-stormy2)*(speed2>10)*(wave2>3)

.
.
. // 2SLS estimates by ethnicity using ivregress
. ivregress 2sls lnqty_a day1 day2 day3 day4 (lnprice_a=speed3)

Instrumental variables (2SLS) regression               Number of obs =      97
                                                       Wald chi2(5)  =   12.84
                                                       Prob > chi2   =  0.0249
                                                       R-squared     =       .
                                                       Root MSE      =  .98121

------------------------------------------------------------------------------
     lnqty_a |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   lnprice_a |  -1.893472    .847839    -2.23   0.026    -3.555206   -.2317384
        day1 |  -.6403019    .321789    -1.99   0.047    -1.270997   -.0096071
        day2 |  -.8393758   .3151283    -2.66   0.008    -1.457016   -.2217356
        day3 |  -.4767465   .3102943    -1.54   0.124    -1.084912    .1314192
        day4 |  -.0592262   .3148365    -0.19   0.851    -.6762943     .557842
       _cons |   7.390611   .3267003    22.62   0.000     6.750291    8.030932
------------------------------------------------------------------------------
Instrumented:  lnprice_a
Instruments:   day1 day2 day3 day4 speed3

. ivregress 2sls lnqty_w day1 day2 day3 day4 (lnprice_w=speed3)

Instrumental variables (2SLS) regression               Number of obs =      97
                                                       Wald chi2(5)  =   29.62
                                                       Prob > chi2   =  0.0000
                                                       R-squared     =  0.1437
                                                       Root MSE      =   .8114

------------------------------------------------------------------------------
     lnqty_w |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   lnprice_w |  -1.029889    .579758    -1.78   0.076    -2.166194    .1064154
        day1 |   .1358369   .2647432     0.51   0.608    -.3830503    .6547241
        day2 |  -.6979756    .260092    -2.68   0.007    -1.207747   -.1882046
        day3 |  -.5011328   .2622948    -1.91   0.056    -1.015221    .0129556
        day4 |   .5185477   .2623988     1.98   0.048     .0042555     1.03284
       _cons |   6.941718   .2250657    30.84   0.000     6.500598    7.382839
------------------------------------------------------------------------------
Instrumented:  lnprice_w
Instruments:   day1 day2 day3 day4 speed3
```

```
. ivregress 2sls lnqty_a day1 day2 day3 day4 (lnprice_a=mixed3 stormy3)

Instrumental variables (2SLS) regression               Number of obs =      97
                                                       Wald chi2(5)  =   16.81
                                                       Prob > chi2   =  0.0049
                                                       R-squared     =       .
                                                       Root MSE      =  .91003

------------------------------------------------------------------------------
     lnqty_a |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   lnprice_a |  -1.564992   .5645862    -2.77   0.006     -2.67156   -.4584233
        day1 |  -.6233142    .297102    -2.10   0.036    -1.205623    -.041005
        day2 |  -.8307524   .2919164    -2.85   0.004    -1.402898   -.2586067
        day3 |  -.4776706   .2877823    -1.66   0.097    -1.041714    .0863724
        day4 |  -.0798918   .2899617    -0.28   0.783    -.6482063    .4884227
       _cons |   7.484395   .2595982    28.83   0.000     6.975591    7.993198
------------------------------------------------------------------------------
Instrumented:  lnprice_a
Instruments:   day1 day2 day3 day4 mixed3 stormy3

. ivregress 2sls lnqty_w day1 day2 day3 day4 (lnprice_w=mixed3 stormy3)

Instrumental variables (2SLS) regression               Number of obs =      97
                                                       Wald chi2(5)  =   30.70
                                                       Prob > chi2   =  0.0000
                                                       R-squared     =  0.1500
                                                       Root MSE      =  .80844

------------------------------------------------------------------------------
     lnqty_w |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   lnprice_w |  -1.007577   .5007281    -2.01   0.044    -1.988986   -.0261676
        day1 |   .1348981   .2634996     0.51   0.609    -.3815515    .6513478
        day2 |  -.6976338   .2591062    -2.69   0.007    -1.205473   -.1897951
        day3 |  -.5032274   .2599362    -1.94   0.053    -1.012693    .0062382
        day4 |   .5164339   .2600145     1.99   0.047     .0068148    1.026053
       _cons |   6.946844    .214264    32.42   0.000     6.526894    7.366794
------------------------------------------------------------------------------
Instrumented:  lnprice_w
Instruments:   day1 day2 day3 day4 mixed3 stormy3

.
. // 2SLS estimates by ethnicity using ivregress - with HAC s.e.s
. tsset t
        time variable:  t, 1 to 97
                delta:  1 unit

. ivregress 2sls lnqty_a day1 day2 day3 day4 (lnprice_a=mixed3
stormy3),vce(hac nw
> est opt)

Instrumental variables (2SLS) regression               Number of obs =      97
                                                       Wald chi2(5)  =   42.56
                                                       Prob > chi2   =  0.0000
```