Nonlinear Statistical Models, Chapter 6,
Nonlinear Simultaneous Equations Models

by

A. Ronald Gallant
Department of Economics
University of North Carolina
Chapel Hill NC 27599-3305 USA

References

Gallant, A. Ronald (1987) *Nonlinear Statistical Models,* Wiley, New York.

Topics

- The Simultaneous Equations Model
- Nonlinear Three Stage Least Squares
  - Example
  - Method
- Generalized Method of Moments
  - Example
  - Method
- Hypothesis Tests

Statistical Model

$$\left.\begin{array}{rcl} q_1(y,x,\theta) & = & e_1 \\ q_2(y,x,\theta) & = & e_2 \\ & \vdots & \\ q_M(y,x,\theta) & = & e_M \end{array}\right\} M \text{ equations}$$

or

$$q(y,x,\theta) = e$$

when written as a vector, where

y	the dependent variables	$L \times 1$
x	the explanatory variables	$k \times 1$
θ	model parameters	$p \times 1$
e	errors	$M \times 1$

If $L > M$, it means that some equations from the complete system are missing.

Data and Error Assumptions

Observed data: $y_t, x_t \quad t = 1, \ldots, n$

Errors: $e_t = q(y_t, x_t, \theta^o)$

First moment: $\mathcal{E} e_t = 0$

Second moment:

Two-Stage Least Squares: iid with

$$\mathcal{E} e_t e_t' = \Sigma \quad \mathcal{E} e_s x_t' = 0$$

or independent with

$$\mathcal{E} e_t e_t' = \Sigma_t \quad \mathcal{E} e_s x_t' = 0$$

Generalized Method of Moments: (x_t, e_t) satisfies mixing conditions. Interest focuses on transformations of x_t of the form $z_t = Z(x_t)$ and the moments

$$\mathcal{E}(e_s \otimes z_s)(e_t \otimes z_t)' = \Sigma_{s,t}$$

There will often be lagged values of y_t among the x_t.

Topics

- The Simultaneous Equations Model
- Nonlinear Three Stage Least Squares
  - Example
  - Method
- Generalized Method of Moments
  - Example
  - Method
- Hypothesis Tests

Table 1a. Household Electricity Expenditures by Time-of-Use, North Carolina, Average over Weekdays in July 1978.

		Expenditure Share			Expenditure
t	Treatment	Base	Intermediate	Peak	($ per day)
1	1	0.056731	0.280382	0.662888	0.46931
2	1	0.103444	0.252128	0.644427	0.79539
3	1	0.158353	0.270089	0.571558	0.45756
4	1	0.108075	0.305072	0.586853	0.94713
5	1	0.083921	0.211656	0.704423	1.22054
6	1	0.112165	0.290532	0.597302	0.93181
7	1	0.071274	0.240518	0.688208	1.79152
8	1	0.076510	0.210503	0.712987	0.51442
9	1	0.066173	0.202999	0.730828	0.78407
10	1	0.094836	0.270281	0.634883	1.01354
11	1	0.078501	0.293953	0.627546	0.83854
12	1	0.059530	0.228752	0.711718	1.53957
13	1	0.208982	0.328053	0.462965	1.06694
14	1	0.083702	0.297272	0.619027	0.82437
15	1	0.138705	0.358329	0.502966	0.80712
16	1	0.111378	0.322564	0.566058	0.53169
17	1	0.092919	0.259633	0.647448	0.85439
18	1	0.039353	0.158205	0.802442	1.93326
19	1	0.066577	0.247454	0.685970	1.37160
20	2	0.102844	0.244335	0.652821	0.92766
21	2	0.125485	0.230305	0.644210	1.80934
22	2	0.154316	0.235135	0.610549	2.41501
.					
.					
.					

Source: Gallant (1987);
Files: electric.doc, electa.dat

Table 1b. Experimental Rates in Effect on a Weekday in July 1978.

	Price (cents per kwh)		
Treatment	Base	Intermedite	Peak
1	1.06	2.86	3.90
2	1.78	2.86	3.90
3	1.06	3.90	3.90
4	1.78	3.90	3.90
5	1.37	3.34	5.06
6	1.06	2.86	6.56
7	1.78	2.86	6.56
8	1.06	3.90	6.56
9	1.78	3.90	6.56

Base period hours are 11pm to 7am. Intermediate period hours are 7am to 10am and 8pm to 11pm. Peak period hours are 10am to 8pm.

Source: Gallant (1987);
Files: electric.doc, electb.dat

Table 1c. Consumer Demographic Characteristics (1)

t	Family Size	Income ($ per yr)	Residence: Size (SqFt)	Residence: Heat Loss (Btuh)
1	2	17000	600	4305
2	6	13500	900	9931
3	2	7000	1248	18878
4	3	11000	1787	17377
5	4	27500	2900	24894
6	3	13500	2000	22526
7	4	22500	3800	17335
8	7	3060	216	4496
9	3	7000	1000	8792
10	1	6793	1200	14663
11	5	11000	1000	14480
12	5	17000	704	3192
13	3	5500	2100	8631
14	2	13500	1400	19720
15	4	22500	1252	7386
16	7	17000	916	7194
17	2	11000	1800	17957
18	2	13500	780	4641
19	3	6570	960	11396
20	4	9000	768	8195
21	2	11000	1200	7812
22	4	13500	900	8878
.				
.				
.				

Source: Gallant (1987);
Files: electric.doc, electc1.dat, electc2.dat

Table 1c. Consumer Demographic Characteristics (2)

t	Elec. Range (1=yes)	Washer (1=yes)	Dryer (1=yes)	Air Condition.: Central (1=yes)	Air Condition.: Window (Btuh)
1	0	1	0	0	13000
2	1	1	0	0	0
3	1	1	0	0	0
4	1	1	0	0	0
5	1	0	0	1	5000
6	1	1	1	0	24000
7	1	1	1	1	0
8	1	0	0	0	0
9	0	1	1	0	18000
10	0	0	0	0	.
11	1	1	0	0	0
12	1	1	1	1	24000
13	1	1	0	1	0
14	1	1	1	0	19000
15	1	1	1	0	24000
16	0	1	0	0	0
17	1	1	1	1	0
18	1	1	0	1	0
19	1	1	0	0	24000
20	1	1	1	0	0
21	1	1	1	1	10000
22	1	1	1	1	0
.					
.					
.					

Source: Gallant (1987);
Files: electric.doc, electc1.dat, electc2.dat

SAS code (data preparation) (1)

```
data raw;
  infile "electa.dat";
  input t1 treat base inter peak expend;
  infile "electc1.dat";
  input t2 famsize income sqfeet heatlos range wash dry cac wac;
  infile "electc2.dat";
  input t3 single duplex mobile hwh frez ref;
  t=t1;
  if (t ne t2) or (t ne t3) then put 'error reading data ' t1 t2 t3;
  drop t1 t2 t3;
```


SAS code (data preparation) (2)

```
data electric;
  set raw;
  if treat=1 then do;  p1=3.90; p2=2.86; p3=1.06;  end;
  if treat=2 then do;  p1=3.90; p2=2.86; p3=1.78;  end;
  if treat=3 then do;  p1=3.90; p2=3.90; p3=1.06;  end;
  if treat=4 then do;  p1=3.90; p2=3.90; p3=1.78;  end;
  if treat=5 then do;  p1=5.06; p2=3.34; p3=1.37;  end;
  if treat=6 then do;  p1=6.56; p2=2.86; p3=1.06;  end;
  if treat=7 then do;  p1=6.56; p2=2.86; p3=1.78;  end;
  if treat=8 then do;  p1=6.56; p2=3.90; p3=1.06;  end;
  if treat=9 then do;  p1=6.56; p2=3.90; p3=1.78;  end;
  y1=log(peak/base);
  y2=log(inter/base);
  y3=log(expend);
  x1=log(p1/expend);
  x2=log(p2/expend);
  x3=log(p3/expend);
  r1=log(p1);
  r2=log(p2);
  r3=log(p3);
  d0=1;
  d1=log((10*p1+6*p2+8*p3)/24);
  d2=log(income);
  d3=log(sqfeet);
  d4=duplex;
  d5=mobile;
  d6=cac*log(heatlos);
  d7=0; if wac>0 then d7=log(wac);
  d8=0; if hwh>0 then d8=log(famsize+1);
  d9=0; if (hwh>0) & (wash>0) then d9=1;
  d10=0; if dry>0 then d10=log(famsize+1);
  d11=0; if ref>0 then d11=log(ref);
  d12=0; if frez>0 then d12=log(frez);
  d13=range;
  keep y1 y2 y3 x1 x2 x3 r1 r2 r3
    d0 d1 d2 d3 d4 d5 d6 d7 d8 d9 d10 d11 d12 d13;
```

Example 1, Chapter 6, NLSM (1)

Model variables:

$$\begin{aligned}
y_1 &= \log\left(\frac{\text{peak expenditure}}{\text{base expenditure}}\right) \\
y_2 &= \log\left(\frac{\text{intermediate expenditure}}{\text{base expenditure}}\right) \\
y_3 &= \log(\text{expenditure}) \\
r_1 &= \log(\text{peak price}) \\
r_2 &= \log(\text{intermediate price}) \\
r_3 &= \log(\text{base price}) \\
d_0 &= 1 \\
d_1 &= \log\left(\frac{\text{10 x peak price} + \text{6 x inter price} + \text{8 x base price}}{24}\right) \\
d_2 &= \log(\text{income})
\end{aligned}$$

- •
- •
- •

Example 1, Chapter 6, NLSM

Complete system:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{pmatrix} = \begin{pmatrix} \log \frac{a_1 + r_t' b_{(1)} - y_{3t} 1' b_{(1)}}{a_3 + r_t' b_{(3)} - y_{3t} 1' b_{(3)}} \\ \log \frac{a_2 + r_t' b_{(2)} - y_{3t} 1' b_{(2)}}{a_3 + r_t' b_{(3)} - y_{3t} 1' b_{(3)}} \\ d'c \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \\ e_{3t} \end{pmatrix}$$

Equations of interest:

$$q(y, x, \theta) = \begin{pmatrix} y_1 - \log \frac{a_1 + r' b_{(1)} - y_3 1' b_{(1)}}{a_3 + r' b_{(3)} - y_3 1' b_{(3)}} \\ y_2 - \log \frac{a_2 + r' b_{(2)} - y_3 1' b_{(2)}}{a_3 + r' b_{(3)} - y_3 1' b_{(3)}} \end{pmatrix}$$

Parameters of interest: (symmetry imposed)

$$a = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \qquad b = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{12} & b_{22} & b_{23} \\ b_{13} & b_{23} & b_{33} \end{pmatrix} = \begin{pmatrix} b'_{(1)} \\ b'_{(2)} \\ b'_{(3)} \end{pmatrix}$$

$$\theta = (a_1, b_{11}, b_{12}, b_{13}, a_2, b_{22}, b_{33})' \quad a_3 = -1$$

Three Stage Least Squares Estimators

Model:

$$q(y_t, x_t, \theta) = e_t \quad M \times 1$$

Instrumental variables:

$$z_t = Z(x_t) \quad K \times 1$$

Moment equations:

$$m_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} m(y_t, x_t, \theta) \quad MK \times 1$$

where

$$\begin{aligned}
m(y_t, x_t, \theta) &= q(y_t, x_t, \theta) \otimes z_t \qquad MK \times 1 \\
&= \begin{pmatrix} q_1(y_t, x_t, \theta) \cdot z_t \\ q_2(y_t, x_t, \theta) \cdot z_t \\ \vdots \\ q_M(y_t, x_t, \theta) \cdot z_t \end{pmatrix} \quad MK \times 1
\end{aligned}$$

Method of Moments Estimator

In classical method of moments, the estimator $\hat{\theta}$ is obtained by setting sample moments to their expectation and solving the resulting equations. The sample moments are

$$m_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} m(y_t, x_t, \theta)$$

the population moments are

$$\mathcal{E} m_n(\theta^o) = \frac{1}{n} \sum_{t=1}^{n} \mathcal{E}(e_t \otimes z_t) = 0$$

The method of moments estimator would then be that $\hat{\theta}$ that solved

$$m_n(\theta) = 0$$

This will only work if $p = MK$ exactly. If $p < MK$, which is the usual case, one uses minimum chi squared instead.

Minimum Chi Squared

The minimum chi squared criterion is

$$S(\theta, W) = n m'_n(\theta) W^{-1} m_n(\theta)$$

where W is a weighting matrix. The optimal choice of W is the variance of $\sqrt{n}\, m_n(\theta^o)$, which is

$$\begin{aligned} W &= n\,\mathcal{E}\Big[m_n(\theta^o) m'_n(\theta^o)\Big] \\ &= \frac{1}{n}\sum_{t=1}^{n} \mathcal{E}(e_t \otimes z_t)(e_t \otimes z_t)' \\ &= \Sigma \otimes \frac{1}{n}\sum_{t=1}^{n} z_t z'_t \end{aligned}$$

The estimator is

$$\hat{\theta}_n = \operatorname*{argmin}_{\theta \in \Theta} S(\theta, W)$$

Generalized Least Squares Analogy

Rather than derive the minimum chi squared estimator from first principles, it is easier show that it is actually generalized least squares in disguise. Put

$$y - f(\theta) = \sqrt{n}\, m_n(\theta).$$

The regression problem is the same as the problem on the previous transparency: There are more equations than unknowns so that the equations $y - f(\theta) = 0$ cannot be solved for θ. When the variance is $\mathcal{E}[y - f(\theta^o)][y - f(\theta^o)]' = \sigma^2 I$, using least squares resolves the problem. When

$$\mathcal{E}[y - f(\theta^o)][y - f(\theta^o)]' = W,$$

using the generalized least squares criterion

$$\begin{aligned} S(\theta, W) &= [y - f(\theta^o)]' W^{-1} [y - f(\theta^o)] \\ &= n\, m'_n(\theta) W^{-1} m_n(\theta) \end{aligned}$$

resolves the problem. This criterion is the same as the chi squared criterion on the previous transparency.

Estimates of W

$$\theta^{\#} = \operatorname*{argmin}_{\theta \in \Theta}\; S\left(\theta, I \otimes \frac{1}{n}\sum_{t=1}^{n} z_t z'_t\right)$$

$$\hat{e}_t = q(y_t, x_t, \hat{\theta}^{\#}) \quad t = 1, \ldots, n$$

Homoskedastic errors:

$$\hat{W} = \left(\frac{1}{n}\sum_{t=1}^{n} \hat{e}_t \hat{e}'_t\right) \otimes \left(\frac{1}{n}\sum_{t=1}^{n} z_t z'_t\right)$$

Heteroskedastic errors:

$$\hat{W} = \frac{1}{n}\sum_{t=1}^{n} (\hat{e}_t \otimes z_t)(\hat{e}_t \otimes z_t)'$$

Asymptotics

$$\sqrt{n}(\hat{\theta}_n - \theta^o) \xrightarrow{\mathcal{L}} N_p(0, V)$$

$$\lim_{n \to \infty} \hat{V}_n = V$$

where

$$\begin{aligned} \hat{V}_n &= \left\{\left(\frac{1}{n}\sum_{t=1}^{n} \hat{Q}_t \otimes z_t\right)' \hat{W}^{-1} \left(\frac{1}{n}\sum_{t=1}^{n} \hat{Q}_t \otimes z_t\right)\right\}^{-1} \\ \hat{Q}_t &= \frac{\partial}{\partial \theta'} q(y_t, x_t, \hat{\theta}_n) \quad M \times p \end{aligned}$$

The easiest way to derive this result is to reuse the generalized least squares analogy $y - f(\theta) = \sqrt{n}\, m_n(\theta)$. The generalized least squares variance estimate is

$$\left(\frac{1}{n}\left\{\frac{\partial}{\partial \theta'}[y - f(\theta)]\right\}' \hat{W}^{-1} \left\{\frac{\partial}{\partial \theta'}[y - f(\theta)]\right\}\right)^{-1},$$

which is the same as $\hat{V}_n$ above.

SAS code (nonlinear three stage least squares)

```
data shat0;
  input _name_ $ y1 y2;
  cards;
  y1  1.0  0.0
  y2  0.0  1.0
  ;

proc model data=electric;
  var y1 y2 y3 r1 r2 r3 d0-d13;
  exogenous r1 r2 r3 d0-d13;
  endogenous y1 y2 y3;
  instruments r1 r2 r3 d0-d13 / noint;
  parms t1=-2.9 t2=-1.3 t3=.82 t4=.36 t5=-1.5 t6=-1. t7=-.03 t8=-.47;
  peak=t1+t2*r1+t3*r2+t4*r3-(t2+t3+t4)*y3;
  inter=t5+t3*r1+t6*r2+t7*r3-(t3+t6+t7)*y3;
  base=-1+t4*r1+t7*r2+t8*r3-(t4+t7+t8)*y3;
  y1=log(peak/base);
  y2=log(inter/base);
  fit y1 y2 / n3sls method=gauss converge=1.e-8 sdata=shat0 outs=shat1;

proc print data=shat1;

proc model data=electric;
  var y1 y2 y3 r1 r2 r3 d0-d13;
  exogenous r1 r2 r3 d0-d13;
  endogenous y1 y2 y3;
  instruments r1 r2 r3 d0-d13 / noint;
  parms t1=-2.9 t2=-1.3 t3=.82 t4=.36 t5=-1.5 t6=-1. t7=-.03 t8=-.47;
  peak=t1+t2*r1+t3*r2+t4*r3-(t2+t3+t4)*y3;
  inter=t5+t3*r1+t6*r2+t7*r3-(t3+t6+t7)*y3;
  base=-1+t4*r1+t7*r2+t8*r3-(t4+t7+t8)*y3;
  y1=log(peak/base);
  y2=log(inter/base);
  fit y1 y2 / n3sls method=gauss converge=1.e-7 sdata=shat1;
```


SAS output (nonlinear three stage least squares)

```
OBS    _NAME_    _TYPE_    _NUSED_       Y1          Y2

 1       Y1       3SLS       220      0.17159    0.096750
 2       Y2       3SLS       220      0.09675    0.095447


             Nonlinear 3SLS Parameter Estimates

                              Approx.        'T'   Approx.
  Parameter    Estimate      Std Err      Ratio  Prob>|T|

  T1          -2.137878      0.58954      -3.63    0.0004
  T2          -1.989392      0.75921      -2.62    0.0094
  T3           0.709391      0.15657       4.53    0.0001
  T4           0.336634      0.05095       6.61    0.0001
  T5          -1.401997      0.15226      -9.21    0.0001
  T6          -1.138897      0.18429      -6.18    0.0001
  T7           0.029131      0.04560       0.64    0.5236
  T8          -0.500502      0.04517     -11.08    0.0001


    Number of Observations       Statistics for System
    Used                220      Objective      0.1589
    Missing               4      Objective*N   34.9640
```


Identification Condition

$$m_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} q(y_t, x_t, \theta) \otimes z_t \qquad MK \times 1$$

Either

$$\theta = \theta^o \Rightarrow \lim_{n\to\infty} m_n(\theta) = 0$$

$$\theta \neq \theta^o \Rightarrow \lim_{n\to\infty} m_n(\theta) \neq 0$$

or

$$\theta = \theta^o \Rightarrow \mathcal{E} m_n(\theta) = 0$$

$$\theta \neq \theta^o \Rightarrow \mathcal{E} m_n(\theta) \neq 0$$

which are equivalent.

Gauss-Newton Downhill Direction

$$\begin{aligned} D(\theta, W) \;=\; & \left\{ \frac{1}{n}\sum_{t=1}^{n} \left[\frac{\partial}{\partial\theta'} q(y_t, x_t, \theta) \otimes z_t\right]' W^{-1} \left[\frac{\partial}{\partial\theta'} q(y_t, x_t, \theta) \otimes z_t\right] \right\}^{-1} \\ & \times \left\{ \frac{1}{n}\sum_{t=1}^{n} \left[\frac{\partial}{\partial\theta'} q(y_t, x_t, \theta) \otimes z_t\right]' W^{-1} [q(y_t, x_t, \theta) \otimes z_t] \right\} \end{aligned}$$

The Modified Gauss-Newton Algorithm

0. Choose a starting value θ_0. Compute

$$D_0 = D(\theta_0, W)$$

Find λ_0 between 0 and 1 such that

$$S(\theta_0 + \lambda_0 D_0, W) < S(\theta_0, W)$$

1. Put $\theta_1 = \theta_0 + \lambda_0 D_0$. Compute

$$D_1 = D(\theta_1, W)$$

Find λ_1 between 0 and 1 such that

$$S(\theta_1 + \lambda_1 D_1, W) < S(\theta_1, W)$$

2. Put $\theta_2 = \theta_1 + \lambda_1 D_1$.

-
-
-

Optimal Instruments

Amemiya (1977) *Econometrica*

$$z_t = \operatorname{vec} \mathcal{E}\left[\frac{\partial}{\partial \theta'} q(y_t, x_t, \theta^o)\right]$$

The result is impractical because one would have to know both θ^o and the error density $p(e|\Sigma)$ to use this result. It would actually be easier to use maximum likelihood (next transparency). There is a literature on trying to estimate z_t for use in N3SLS. Most people just use low order polynomials

$$z_t = (1, x_{1t}, \ldots, x_{kt}, x_{1t}^2, x_{1t}x_{2t}, \ldots, x_{kt}^2, \ldots)$$

Often just to the first order, as in Example 1 above, where $z_t = (1, x_{1t}, \ldots, x_{kt})$.

Maximum Likelihood

$e \sim p(e, \Sigma)$ density for e

$e = q(y, x, \theta)$ the transformation

$\frac{\partial}{\partial y'} q(y, x, \theta)$ the Jacobian

The density for y is

$$p(y|x, \theta, \Sigma) = \left|\det \frac{\partial}{\partial y'} q(y, x, \theta)\right| p[q(y, x, \theta)|\Sigma]$$

and the log likelihood is

$$\ell(\theta, \Sigma) = \sum_{t=1}^{n} \log p(y_t|x_t, \theta, \Sigma)$$

SAS code (maximum likelihood)

```
proc model data=electric;
  var y1 y2 y3 r1 r2 r3 d0-d13;
  exogenous r1 r2 r3 d0-d13;
  endogenous y1 y2 y3;
  parms t1=-2.9 t2=-1.3 t3=.82 t4=.36 t5=-1.5 t6=-1. t7=-.03 t8=-.47
            c0=0 c1=0 c2=0 c3=0 c4=0 c5=0 c6=0 c7=0 c8=0 c9=0 c10=0
            c11=0 c12=0 c13=0;
  peak = t1 + t2*r1 + t3*r2 + t4*r3 - (t2+t3+t4)*y3;
  inter= t5 + t3*r1 + t6*r2 + t7*r3 - (t3+t6+t7)*y3;
  base=  -1 + t4*r1 + t7*r2 + t8*r3 - (t4+t7+t8)*y3;
  y1 = log(peak/base);
  y2 = log(inter/base);
  y3 = c0 + c1*d1 + c2*d2 + c3*d3 + c4*d4 + c5*d5 + c6*d6 + c7*d7
       + c8*d8 + c9*d9 + c10*d10 + c11*d11 + c12*d12 + c13*d13;
  fit y1 y2 y3 / fiml;
```

SAS output (maximum likelihood)

Nonlinear FIML Parameter Estimates

Parameter	Estimate	Approx. Std Err	'T' Ratio	Approx. Prob>|T|
T1	-2.436624	0.35068	-6.95	0.0001
T2	-1.656306	0.34301	-4.83	0.0001
T3	0.766855	0.09224	8.31	0.0001
T4	0.355883	0.03383	10.52	0.0001
T5	-1.490733	0.11093	-13.44	0.0001
T6	-1.043513	0.10141	-10.29	0.0001
T7	0.031660	0.04142	0.76	0.4454
T8	-0.463048	0.02594	-17.85	0.0001
C0	-4.232964	0.74109	-5.71	0.0001
C1	1.223692	0.25305	4.84	0.0001
C2	0.146805	0.06256	2.35	0.0199
C3	0.097988	0.09453	1.04	0.3011
C4	-0.052285	0.18754	-0.28	0.7807
C5	-0.121091	0.12207	-0.99	0.3224
C6	0.069855	0.01093	6.39	0.0001
C7	0.039399	0.0088411	4.46	0.0001
C8	0.270019	0.08873	3.04	0.0026
C9	0.00361939	0.11561	0.03	0.9751
C10	0.065136	0.06161	1.06	0.2916
C11	0.191763	0.08107	2.37	0.0189
C12	0.265636	0.13071	2.03	0.0434
C13	0.108342	0.10780	1.01	0.3161

```
Number of Observations       Statistics for System
Used                 220     Log Likelihood   -207.3811
Missing                4
```


Topics

- The Simultaneous Equations Model
- Nonlinear Three Stage Least Squares
  - Example
  - Method
- Generalized Method of Moments
  - Example
  - Method
- Hypothesis Tests

Table 1a. Consumption and Stock Returns.

t	Year	Month	Nondurables and Services	Population	Value Weighted NYSE Returns	Implicit Deflator
0	1959	1	381.9	176.6850	0.0093695102	0.6818539
1	.	2	383.7	176.9050	0.0093310997	0.6823039
2	.	3	388.3	177.1460	0.0049904501	0.6814319
3	.	4	385.5	177.3650	0.0383739690	0.6830091
4	.	5	389.7	177.5910	0.0204769890	0.6846292
5	.	6	390.0	177.8300	0.0007165600	0.6876923
6	.	7	389.2	178.1010	0.0371922290	0.6893628
7	.	8	390.7	178.3760	-0.0113433900	0.6910673
8	.	9	393.6	178.6570	-0.0472779090	0.6930894
9	.	10	394.2	178.9210	0.0164727200	0.6945713
10	.	11	394.1	179.1530	0.0194594210	0.6950013
11	.	12	396.5	179.3860	0.0296911900	0.6958386
12	1960	1	396.8	179.5970	-0.0664901060	0.6960685
13	.	2	395.4	179.7880	0.0114439700	0.6967628
14	.	3	399.1	180.0070	-0.0114419700	0.6983212
15	.	4	404.2	180.2220	-0.0163223000	0.7013855
16	.	5	399.8	180.4440	0.0328373610	0.7016008
17	.	6	401.3	180.6710	0.0231378990	0.7024670
18	.	7	402.0	180.9450	-0.0210754290	0.7034826
19	.	8	400.4	181.2380	0.0296860300	0.7047952
20	.	9	400.2	181.5280	-0.0568203400	0.7061469
21	.	10	402.9	181.7960	-0.0045937700	0.7078680
22	.	11	403.8	182.0420	0.0472565590	0.7100049
23	.	12	401.6	182.2870	0.0478186380	0.7109064
.						
.						
.						

Source: Gallant (1987);
Files: hansen.doc, hansena.dat, hansenb.dat

SAS code (data preparation)

```
data hansen;
  infile "hansena.dat";
  input t year month nds people stocks deflator;
  ndsper = nds/people;
  y = ndsper/lag(ndsper);
  x = (1 + stocks)*lag(deflator)/deflator;
  z0 = 1;
  z1 = lag(y);
  z2 = lag(x);
  e = 0;
  if ( z1 = . ) then delete;
  keep y x z0 z1 z2 e;
```

Example 2, Chapter 6, NLSM

Model variables:

$$y_t = \frac{\text{per capita consumption at time } t}{\text{per capita consumption at time } t-1}$$
$$x_t = 1 + \text{ real stock returns at time } t$$

Model:

$$q(y_t, x_t, \theta) = \beta x_t y_t^{\alpha} - 1 \quad M = 1 \quad L = 2$$

Instruments:

$$z_t = (1, y_{t-1}, x_{t-1})' \quad K = 3$$

Parameters:

$$\theta = (\alpha, \beta)' \quad p = 2$$

Moment equations:

$$m_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} q(y_t, x_t, \theta) \otimes z_t \quad MK = 3$$

GMM Error Assumptions

$$\mathcal{E}(e_t \otimes z_t) = 0$$
$$\mathcal{E}(e_t \otimes z_t)(e_s \otimes z_s) = \Sigma_{st}$$

and mixing conditions.

This is the heterogeneous, serially correlated case of N3SLS. GMM is the same as N3SLS except that one uses a HAC weighting matrix as implied by the error assumptions.

GMM Weighting Matrix

$$\hat{\theta}^{\#} = \underset{\theta \in \Theta}{\text{argmin}}\, S\left(\theta, I \otimes \frac{1}{n}\sum_{t=1}^{n} z_t z_t'\right)$$
$$\hat{e}_t = q(y_t, x_t, \hat{\theta}^{\#}) \quad t = 1, \ldots, n$$
$$\hat{W} = \sum_{t=l(n)}^{l(n)} w\left(\frac{\tau}{l(n)}\right) \hat{W}_{n\tau}$$

where $l(n) = n^{1/5}$ and

$$\hat{W}_{n\tau} = \begin{cases} \frac{1}{n}\sum_{t=\tau+1}^{n} (\hat{e}_t \otimes z_t)\,(\hat{e}_{t-\tau} \otimes z_{t-\tau})' & \tau \geq 0 \\ \hat{W}'_{n,-\tau} & \tau < 0 \end{cases}$$

$$w(v) = \begin{cases} 1 - 6|v|^2 + 6|v|^3 & 0 \leq |v| \leq \frac{1}{2} \\ 2(1-|v|)^3 & \frac{1}{2} \leq |v| \leq 1 \end{cases}$$

GMM Estimator

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\text{argmin}}\, S(\theta, \hat{W}_n)$$
$$S(\theta, W) = n\, m_n'(\theta) W^{-1} m_n(\theta)$$

Asymptotics (same as N3SLS)

$$\sqrt{n}\,(\hat{\theta}_n - \theta^o) \xrightarrow{\mathcal{L}} N_p(0, V)$$
$$\lim_{n\to\infty} \hat{V}_n = V$$

where

$$\hat{V}_n = \left\{\left(\frac{1}{n}\sum_{t=1}^{n} \hat{Q}_t \otimes z_t\right)' \hat{W}^{-1} \left(\frac{1}{n}\sum_{t=1}^{n} \hat{Q}_t \otimes z_t\right)\right\}^{-1}$$
$$\hat{Q}_t = \frac{\partial}{\partial \theta'} q(y_t, x_t, \hat{\theta}_n) \quad M \times p$$

SAS code (GMM)

```
proc model data=hansen;
  var y x z0 z1 z2 e;
  endogenous y x e;
  exogenous z0 z1 z2;
  instruments z0 z1 z2 / noint;
  parms alpha -0.4 beta 0.9;
  e=beta*(y**alpha)*x-1;
  fit e / gmm method=gauss kernel=(parzen,0,0.2);

* This choice of kernel is to force l(n)=0 because the errors in this
  model should be independent from economic theory.  In general one
  should use kernel=(parzen,1.0,0.2), which is the default.;
```


SAS output (GMM)

```
              Nonlinear GMM Parameter Estimates

                                Approx.      'T'   Approx.
     Parameter    Estimate     Std Err     Ratio  Prob>|T|

     ALPHA       -1.033594     1.90011     -0.54    0.5870
     BETA         0.998257   0.0045583    219.00    0.0001


       Number of Observations       Statistics for System
       Used               238       Objective    0.004404
       Missing              0       Objective*N    1.0480
```


Topics

- The Simultaneous Equations Model
- Nonlinear Three Stage Least Squares
  - Example
  - Method
- Generalized Method of Moments
  - Example
  - Method
- Hypothesis Tests

Simultaneous Systems: Hypothesis Testing (1)

We will illustrate using Example 1

$$q(y_t, r_t, \theta) = \begin{pmatrix} y_{1t} - \log \frac{\theta_1+\theta_2 r_{1t}+\theta_3 r_{2t}+\theta_4 r_{3t}+(\theta_2+\theta_3+\theta_4)y_{3t}}{-1+\theta_4 r_{1t}+\theta_7 r_{2t}+\theta_8 r_{3t}+(\theta_4+\theta_7+\theta_8)y_{3t}} \\ y_{1t} - \log \frac{\theta_5+\theta_3 r_{1t}+\theta_6 r_{2t}+\theta_7 r_{3t}+(\theta_3+\theta_6+\theta_7)y_{3t}}{-1+\theta_4 r_{1t}+\theta_7 r_{2t}+\theta_8 r_{3t}+(\theta_4+\theta_7+\theta_8)y_{3t}} \end{pmatrix}$$

and testing the hypothesis of straight-line Engle curves (linear homogeneity), which is

$$\text{H} : h(\theta^o) = 0 \text{ against A} : h(\theta^o) \neq 0$$

where

$$h(\theta) = \begin{pmatrix} \theta_2 + \theta_3 + \theta_4 \\ \theta_3 + \theta_6 + \theta_7 \\ \theta_4 + \theta_7 + \theta_8 \end{pmatrix}$$

Simultaneous Systems: Hypothesis Testing (2)

There are no new ideas because, as we have seen, estimation based on the moment equations

$$m_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} q(y_t, x_t, \theta) \otimes z_t$$

can be viewed as a generalized least squares problem. The theory differs somewhat, but the algebra works out correctly. The test statistics that result can be unravelled to get expressions for test statistics in the systems notation.

There is nothing enlightening about this exercise. Therefore the expressions shall merely be given without derivations.

Wald Test: Some Preliminary Notation

$$h : \Theta \to R^q \quad \hat{h} = h(\hat{\theta}) \quad \hat{H} = \frac{\partial}{\partial \theta'} h(\hat{\theta})$$

$$\sqrt{n}\,(\hat{\theta}_n - \theta^o) \xrightarrow{\mathcal{L}} N_p(0, V)$$

$$\hat{V}_n = \left\{ \left(\frac{1}{n}\sum_{t=1}^{n} \hat{Q}_t \otimes z_t \right)' \hat{W}^{-1} \left(\frac{1}{n}\sum_{t=1}^{n} \hat{Q}_t \otimes z_t \right) \right\}^{-1}$$

$$\hat{Q}_t = \frac{\partial}{\partial \theta'} q(y_t, x_t, \hat{\theta}_n) \qquad M \times p$$

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\text{argmin}}\; S(\theta, \hat{W})$$

$$S(\theta, W) = n\, m_n(\theta) W^{-1} m_n(\theta)$$

$$m_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} q(y_t, x_t, \theta) \otimes z_t \qquad MK \times 1$$

$$\hat{W} = \widehat{\text{Var}}[\sqrt{n}\, m_n(\theta^o)] \qquad MK \times MK$$

$\hat{W}$ has three possible forms: iid, HI, or HAC.

Wald Test

The Wald test statistic for

$$\text{H} : h(\theta^o) = 0 \text{ against } A : h(\theta^o) \neq 0;$$

is

$$W = n\, \hat{h}' (\hat{H} V \hat{H}')^{-1} \hat{h}$$

Reject H when W exceeds the upper critical point of a $\mathcal{X}^2$ on q degrees of freedom.

Wald Test (SAS code)

```
proc model data=electric;
  var y1 y2 y3 r1 r2 r3 d0-d13;
  exogenous r1 r2 r3 d0-d13;
  endogenous y1 y2 y3;
  instruments r1 r2 r3 d0-d13 / noint;
  parms t1=-2.9 t2=-1.3 t3=.82 t4=.36 t5=-1.5 t6=-1. t7=-.03 t8=-.47;
  peak=t1+t2*r1+t3*r2+t4*r3-(t2+t3+t4)*y3;
  inter=t5+t3*r1+t6*r2+t7*r3-(t3+t6+t7)*y3;
  base=-1+t4*r1+t7*r2+t8*r3-(t4+t7+t8)*y3;
  y1=log(peak/base);
  y2=log(inter/base);
  fit y1 y2 / n3sls method=gauss converge=1.e-7 sdata=shat1;
  test t2+t3+t4=0, t3+t6+t7=0, t4+t7+t8=0 ,/ wald;
```

Wald Test (SAS output)

Test Results

Test	Type	Statistic	Prob.	Label
Test0	Wald	3.01	0.3897	

Constrained and Unconstrained Estimates

$$q(y_t, x_t, \theta^o) = e_t \quad t = 1, \ldots, n$$

$$\text{H} : h(\theta^o) = 0 \text{ against A} : h(\theta^o) \neq 0$$

Unconstrained Estimate:

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\text{argmin}}\; S(\theta, \hat{W})$$

Constrained Estimate:

$$\tilde{\theta}_n = \underset{h(\theta)=0}{\text{argmin}}\; S(\theta, \hat{W})$$

Notice that $\hat{W}$ is the same for both.

Likelihood Ratio Test

aka Gallant-Jorgensen test
J. of Econometrics (1979)

The statistic

$$L = S(\tilde{\theta}_n, \hat{W}) - S(\hat{\theta}_n, \hat{W})$$

is the "likelihood ratio" test statistic for H : $h(\theta^o) = 0$ against $A : h(\theta^o) \neq 0$. **Note that $\hat{W}$ must be the same in both criterion function computations.** L is to be compared to the quantiles of the chi squared distribution on q degrees of freedom. One rejects for large L.

L.R. Test (SAS code)

```
proc model data=electric;
  var y1 y2 y3 r1 r2 r3 d0-d13;
  exogenous r1 r2 r3 d0-d13;
  endogenous y1 y2 y3;
  instruments r1 r2 r3 d0-d13 / noint;
  parms t1=-2.9 t2=-1.3 t3=.82 t4=.36 t5=-1.5 t6=-1. t7=-.03 t8=-.47;
  peak=t1+t2*r1+t3*r2+t4*r3-(t2+t3+t4)*y3;
  inter=t5+t3*r1+t6*r2+t7*r3-(t3+t6+t7)*y3;
  base=-1+t4*r1+t7*r2+t8*r3-(t4+t7+t8)*y3;
  y1=log(peak/base);
  y2=log(inter/base);
  fit y1 y2 / n3sls method=gauss converge=1.e-7 sdata=shat1;
  test t2+t3+t4=0, t3+t6+t7=0, t4+t7+t8=0 ,/ lr;
```

L.R. Test (SAS output)

```
                    Test Results

  Test      Type      Statistic     Prob.    Label

  Test0     L.R.           3.38     0.3361
```

Warning: These results are from SAS Version 8; earlier releases fail.

L.M. Test: Some Preliminary Notation

$$D(\theta, W) = \left\{ \frac{1}{n}\sum_{t=1}^{n} \left[\frac{\partial}{\partial\theta'} q(y_t, x_t, \theta) \otimes z_t\right]' W^{-1} \left[\frac{\partial}{\partial\theta'} q(y_t, x_t, \theta) \otimes z_t\right] \right\}^{-1}$$
$$\times \left\{ \frac{1}{n}\sum_{t=1}^{n} \left[\frac{\partial}{\partial\theta'} q(y_t, x_t, \theta) \otimes z_t\right]' W^{-1} [q(y_t, x_t, \theta) \otimes z_t] \right\}$$

$$\tilde{V}_n^* = \left\{ \left(\frac{1}{n} \sum_{t=1}^{n} \tilde{Q}_t^* \otimes z_t \right)' (\tilde{W}^*)^{-1} \left(\frac{1}{n} \sum_{t=1}^{n} \tilde{Q}_t^* \otimes z_t \right) \right\}^{-1}$$

$$\tilde{Q}_t^* = \frac{\partial}{\partial\theta'} q(y_t, x_t, \tilde{\theta}_n^*) \quad M \times p$$

$$\tilde{\theta}_n^* = \underset{\theta \in \Theta}{\text{argmin}}\; S(\theta, \tilde{W}^*)$$

$$\tilde{\theta}_n^{\#} = \underset{h(\theta)=0}{\text{argmin}}\; S\Big(\theta, I \otimes \frac{1}{n} \sum_{t=1}^{n} z_t z_t'\Big)$$

$$\tilde{e}_t^{\#} = q(y_t, x_t, \tilde{\theta}_n^{\#})$$

$$S(\theta, W) = n\, m_n(\theta) W^{-1} m_n(\theta)$$

$$m_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} q(y_t, x_t, \theta) \otimes z_t \quad MK \times 1$$

$$\tilde{W}^* = \widehat{\text{Var}}[\sqrt{n}\, m_n(\theta^o)] \quad MK \times MK$$

$\tilde{W}^*$ is computed from $\tilde{e}_t^{\#}$ using the appropriate form: iid, HI, or HAC. **Note that $\tilde{\theta}_n^*$ is not $\tilde{\theta}_n$ as defined above.**

Lagrange Multiplier Test

The statistic

$$R = n\, D'(\tilde{\theta}^*, \tilde{W}^*)(\tilde{V}^*)^{-1} D(\tilde{\theta}^*, \tilde{W}^*)$$

is to be compared to the quantiles of the chi squared distribution on q degrees of freedom. One rejects for large R.

L.M. Test (SAS code)

```
proc model data=electric;
  var y1 y2 y3 r1 r2 r3 d0-d13;
  exogenous r1 r2 r3 d0-d13;
  endogenous y1 y2 y3;
  instruments r1 r2 r3 d0-d13 / noint;
  parms t1=-2.9 t2=-1.3 t3=.82 t4=.36 t5=-1.5 t6=-1. t7=-.03 t8=-.47;
  peak=t1+t2*r1+t3*r2+t4*r3-(t2+t3+t4)*y3;
  inter=t5+t3*r1+t6*r2+t7*r3-(t3+t6+t7)*y3;
  base=-1+t4*r1+t7*r2+t8*r3-(t4+t7+t8)*y3;
  y1=log(peak/base);
  y2=log(inter/base);
  fit y1 y2 / n3sls method=gauss converge=1.e-7 sdata=shat1;
  test t2+t3+t4=0, t3+t6+t7=0, t4+t7+t8=0 ,/ lm;
```

L.M. Test (SAS output)

```
                     Test Results

  Test       Type        Statistic      Prob.    Label

  Test0      L.M.              .          .
```

Warning: SAS Version 8 or higher required; earlier releases fail as above.

Test of Overidentifying Restrictions

$$\text{H} : \mathcal{E}[m_n(\theta)] = 0 \text{ for some } \theta$$

against

$$\text{A} : \mathcal{E}[m_n(\theta)] \neq 0 \text{ for any } \theta$$

Under the null hypothesis,

$$J = S(\hat{\theta}_n, \hat{W})$$

is distributed as a chi squared with $MK - p$ degrees of freedom.

This is a standard result from minimum chi squared estimation. It can be seen from the regression analogy. Viewing $\sqrt{n}\, m_n(\theta)$ as the equations defining a nonlinear generalized least squares regression, J is the SSE for error. But we have essentially assumed that we know W exactly, not just to within a scalar; that is, we assumed that Var$(\sqrt{n}\, m_n(\theta))$ is equal to W rather than $\sigma^2 W$. Therefore, $J/(MK-p)$ is estimating 1. If $J/(MK-p)$ is markedly bigger than 1 as measured by the chi squared critical point, the null hypothesis is suspect.

Confidence Intervals

There is nothing special about the simultaneous equations case. The ideas are the same as before:

To set a confidence interval on a nonlinear function $\gamma(\theta)$, invert one of the three tests. That is, let

$$h(\theta) = \gamma(\theta) - \gamma^*$$

and put in the interval all γ^* for which

$$\text{H} : h(\theta) = 0$$

is accepted.

Topics

- The Simultaneous Equations Model
- Nonlinear Three Stage Least Squares
  - Example
  - Method
- Generalized Method of Moments
  - Example
  - Method
- Hypothesis Tests