Nonlinear Statistical Models, Chapter 5, Multivariate Nonlinear Regression

by

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

© 2000 by A. Ronald Gallant

References

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

Topics

- Least Squares Estimates
- Hypothesis Tests
- Confidence Intervals
- HI and HAC Variance Estimates
- Maximum Likelihood Estimators

Statistical Model

$$y_t = f(x_t, \theta) + e_t \quad t = 1, 2, \ldots, n$$

y_t the dependent variable, M-variate, observed

x_t the explanatory variables, k-variate, observed

θ model parameters, p-variate, unknown (to be estimated)

e_t the error, M-variate, unobserved (because θ is unknown) $\mathcal{E}(e_t) = 0$, $\mathcal{E}(e_t e_t') = \Sigma$, iid

Generalized Least Squares Estimator

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\text{argmin}}\ \text{SSE}(\theta, \Sigma)$$

$$\text{SSE}(\theta, \Sigma) = \sum_{t=1}^{n} [y_t - f(x_t, \theta)]' \, \Sigma^{-1} \, [y_t - f(x_t, \theta)]$$

Derivations

There are two ways to derive computational methods and statistical properties. The first is to derive them afresh from the expression

$$\text{SSE}(\theta,\Sigma) = \sum_{t=1}^{n} [y_t - f(x_t,\theta)]' \, \Sigma^{-1} \, [y_t - f(x_t,\theta)]$$

The second is to rotate to the univariate case and apply previous results. The rotation is

$$\begin{aligned}
\text{SSE}(\theta,\Sigma) &= \sum_{t=1}^{n} [Py_t - Pf(x_t,\theta)]' [Py_t - Pf(x_t,\theta)] \\
&= \sum_{t=1}^{n} \sum_{\alpha=1}^{M} \left[p'_{(\alpha)} y_t - p'_{(\alpha)} f(x_t,\theta)\right]^2 \\
&= \sum_{s=1}^{nM} [\,\text{“}y\text{”}_s - \text{“}f\text{”}\,(\text{“}x\text{”}_s,\theta)]^2
\end{aligned}$$

where

$$\Sigma^{-1} = P'P \quad P = \begin{pmatrix} p'_{(1)} \\ \vdots \\ p'_{(M)} \end{pmatrix} \quad s = M(t-1)+\alpha$$

$$\text{“}y\text{”}_s = p'_{(\alpha)} y_t \, \text{“}f\text{”}\,(\text{“}x\text{”}_s,\theta) = p'_{(\alpha)} f(x_t,\theta) \, \text{“}x\text{”}_s = (p_{(\alpha)}, x_t)$$

With either approach, there are no new ideas involved. Therefore, we shall just present results without detailed derivations.

Estimation of Σ

Compute the nonlinear least square estimator, which is

$$\hat{\theta}_n^{\#} = \underset{\theta\in\Theta}{\text{argmin}} \ \text{SSE}(\theta, I),$$

and put

$$\hat{\Sigma} = \frac{1}{n} \sum_{t=1}^{n} \hat{e}_t \, \hat{e}_t'$$

where

$$\hat{e}_t = y_t - f(x_t, \hat{\theta}^{\#})$$

Summary

$$\text{SSE}(\theta,\Sigma) = \sum_{t=1}^{n} [y_t - f(x_t,\theta)]' \, \Sigma^{-1} \, [y_t - f(x_t,\theta)]$$

$$\hat{\theta}_n^{\#} = \underset{\theta\in\Theta}{\text{argmin}} \ \text{SSE}(\theta, I),$$

$$\hat{\Sigma} = \frac{1}{n} \sum_{t=1}^{n} \left[y_t - f(x_t,\hat{\theta}^{\#})\right] \left[y_t - f(x_t,\hat{\theta}^{\#})\right]'$$

$$\hat{\theta}_n = \underset{\theta\in\Theta}{\text{argmin}} \ \text{SSE}(\theta, \hat{\Sigma})$$

Statistical Properties

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

$$\hat{V} = \left\{ \frac{1}{n} \sum_{t=1}^{n} \left[\frac{\partial}{\partial\theta'} f(x_t,\hat{\theta})\right]' \hat{\Sigma}^{-1} \left[\frac{\partial}{\partial\theta'} f(x_t,\hat{\theta})\right] \right\}^{-1}$$

Gauss-Newton Downhill Direction

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

The Modified Gauss-Newton Algorithm

0. Choose a starting value θ_0. Compute

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

Find λ_0 between 0 and 1 such that

$$\mathrm{SSE}(\theta_0 + \lambda_0 D_0, \Sigma) < \mathrm{SSE}(\theta_0, \Sigma)$$

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

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

Find λ_1 between 0 and 1 such that

$$\mathrm{SSE}(\theta_1 + \lambda_1 D_1, \Sigma) < \mathrm{SSE}(\theta_1, \Sigma)$$

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

-
-
-

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

t	Treatment	Expenditure Share: Base	Expenditure Share: Intermediate	Expenditure Share: Peak	Expenditure ($ 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.

Treatment	Price (cents per kwh): Base	Price (cents per kwh): Intermedite	Price (cents per kwh): 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)

				Air Condition.	
t	Elec. Range (1=yes)	Washer (1=yes)	Dryer (1=yes)	Central (1=yes)	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

Example 1, Chapter 5, NLSM

$$y_t = f(x_t, \theta^o) + e_t \quad t = 1, \dots, 224 = n$$

where

$$f(x,\theta) = \begin{pmatrix} \log \frac{\theta_1 + \theta_2 x_{1t} + \theta_3 x_{2t} + \theta_4 x_{3t}}{-1 + \theta_4 x_{1t} + \theta_7 x_{2t} + \theta_8 x_{3t}} \\ \log \frac{\theta_5 + \theta_3 x_{1t} + \theta_6 x_{2t} + \theta_7 x_{3t}}{-1 + \theta_4 x_{1t} + \theta_7 x_{2t} + \theta_8 x_{3t}} \end{pmatrix}$$

and

$$\begin{aligned} y_1 &= \log\left(\frac{\text{peak expenditure}}{\text{base expenditure}}\right) \\ y_2 &= \log\left(\frac{\text{intermediate expenditure}}{\text{base expenditure}}\right) \\ x_1 &= \log\left(\frac{\text{peak price}}{\text{expenditure}}\right) \\ x_2 &= \log\left(\frac{\text{intermediate price}}{\text{expenditure}}\right) \\ x_3 &= \log\left(\frac{\text{base price}}{\text{expenditure}}\right) \\ \theta &= (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8)' \end{aligned}$$

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;
```

SAS code (multivariate nonlinear regression)

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

proc model data=electric;
  var y1 y2 x1 x2 x3;
  parms t1=-2.9 t2=-1.3 t3=.82 t4=.36 t5=-1.5 t6=-1. t7=-.03 t8=-.47;
  peak=t1+t2*x1+t3*x2+t4*x3;
  inter=t5+t3*x1+t6*x2+t7*x3;
  base=-1+t4*x1+t7*x2+t8*x3;
  y1=log(peak/base);
  y2=log(inter/base);
  fit y1 y2 / sur method=gauss converge=1.e-8 sdata=shat0 outs=shat1;

data shat1;
  set shat1;
  y1=(220.0/224.0)*y1;
  y2=(220.0/224.0)*y2;
* This step is irrelevant, it is here to permit comparison with p.325
  of NLSM;

proc print data=shat1;

proc model data=electric;
  parms t1=-2.9 t2=-1.3 t3=.82 t4=.36 t5=-1.5 t6=-1. t7=-.03 t8=-.47;
  var y1 y2 x1 x2 x3;
  peak=t1+t2*x1+t3*x2+t4*x3;
  inter=t5+t3*x1+t6*x2+t7*x3;
  base=-1+t4*x1+t7*x2+t8*x3;
  y1=log(peak/base);
  y2=log(inter/base);
  fit y1 y2 / sur method=gauss converge=1.e-8 sdata=shat1;
```


SAS output (multivariate nonlinear regression)

```
 OBS    _NAME_    _TYPE_    _NUSED_       Y1         Y2

  1       Y1       SUR        224      0.16492    0.092006
  2       Y2       SUR        224      0.09201    0.089643

                      MODEL Procedure

        Nonlinear SUR Summary of Residual Errors

            DF    DF
 Equation Model Error       SSE       MSE R-Square Adj R-Sq

 Y1           4   220   36.9912   0.16814   0.4717   0.4645
 Y2           4   220   20.1332   0.09151   0.4515   0.4440


           Nonlinear SUR Parameter Estimates

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

  T1          -2.924581      0.27577    -10.61    0.0001
  T2          -1.286746      0.22497     -5.72    0.0001
  T3           0.818570      0.08026     10.20    0.0001
  T4           0.361158      0.03006     12.02    0.0001
  T5          -1.537589      0.09122    -16.86    0.0001
  T6          -1.048959      0.08303    -12.63    0.0001
  T7           0.030087      0.03586      0.84    0.4024
  T8          -0.467420      0.01911    -24.46    0.0001


    Number of Observations       Statistics for System
    Used                 224     Objective      1.9949
    Missing                0     Objective*N  446.8570
```


Topics

- Least Squares Estimates
- Hypothesis Tests
- Confidence Intervals
- HI and HAC Variance Estimates
- Maximum Likelihood Estimators

Multivariate Models: Hypothesis Testing (1)

We will illustrate using Example 1

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \log \frac{\theta_1+\theta_2 x_{1t}+\theta_3 x_{2t}+\theta_4 x_{3t}}{-1+\theta_4 x_{1t}+\theta_7 x_{2t}+\theta_8 x_{3t}} \\ \log \frac{\theta_5+\theta_3 x_{1t}+\theta_6 x_{2t}+\theta_7 x_{3t}}{-1+\theta_4 x_{1t}+\theta_7 x_{2t}+\theta_8 x_{3t}} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}$$

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

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

Multivariate Models: Hypothesis Testing (2)

There are no new ideas because multivariate nonlinear least squares can be rotated to the univariate situation. The univariate test statistic that results can be unravelled to get expressions for test statistics in a multivariate notation.

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

Wald Test: Some Preliminary Notation

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

$$\hat{V} = \left\{ \frac{1}{n} \sum_{t=1}^{n} \left[\frac{\partial}{\partial\theta'} f(x_t, \hat{\theta}) \right]' \hat{\Sigma}^{-1} \left[\frac{\partial}{\partial\theta'} f(x_t, \hat{\theta}) \right] \right\}^{-1}$$

$$\hat{\Sigma} = \frac{1}{n} \sum_{t=1}^{n} \left[y_t - f(x_t, \hat{\theta}^{\#}) \right] \left[y_t - f(x_t, \hat{\theta}^{\#}) \right]'$$

$$\hat{\theta}_n = \underset{\theta\in\Theta}{\text{argmin}}\ \text{SSE}(\theta, \hat{\Sigma})$$

$$\hat{\theta}_n^{\#} = \underset{\theta\in\Theta}{\text{argmin}}\ \text{SSE}(\theta, I)$$

$$\text{SSE}(\theta, \Sigma) = \sum_{t=1}^{n} \left[y_t - f(x_t, \theta) \right]' \Sigma^{-1} \left[y_t - f(x_t, \theta) \right]$$

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.

Or, compare

$$W = \frac{n\,\hat{h}'(\hat{H} V \hat{H}')^{-1} \hat{h} \,/\, q}{\text{SSE}(\hat{\theta}) \,/\, (nM - p)}$$

to the upper critical point of the F-distribution with q numerator degrees of freedom and $nM-p$ denominator degrees of freedom.

Wald Test (SAS code)

```
proc model data=electric;
  parms t1=-2.9 t2=-1.3 t3=.82 t4=.36 t5=-1.5 t6=-1. t7=-.03 t8=-.47;
  var y1 y2 x1 x2 x3;
  peak=t1+t2*x1+t3*x2+t4*x3;
  inter=t5+t3*x1+t6*x2+t7*x3;
  base=-1+t4*x1+t7*x2+t8*x3;
  y1=log(peak/base);
  y2=log(inter/base);
  fit y1 y2 / sur method=gauss converge=1.e-8 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            21.94      0.0001
```

Constrained and Unconstrained Estimates

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

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

Unconstrained Estimate:

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

Constrained Estimate:

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

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

Likelihood Ratio Test

The statistic

$$L = \text{SSE}(\tilde{\theta}_n, \hat{\Sigma}) - \text{SSE}(\hat{\theta}_n, \hat{\Sigma})$$

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

Often one computes

$$L = \frac{\left[\text{SSE}(\tilde{\theta}_n, \hat{\Sigma}) - \text{SSE}(\hat{\theta}_n, \hat{\Sigma})\right]/q}{\text{SSE}(\hat{\theta}_n, \hat{\Sigma})/(nM - p)}$$

instead and compares to the quantiles of the F-distribution with q numerator degrees of freedom and $nM - p$ denominator degrees of freedom because this agrees with the formulas used in linear models and gives more accurate answers in small samples.

L.R. Test (SAS code)

```
proc model data=electric;
  parms t1=-2.9 t2=-1.3 t3=.82 t4=.36 t5=-1.5 t6=-1. t7=-.03 t8=-.47;
  var y1 y2 x1 x2 x3;
  peak=t1+t2*x1+t3*x2+t4*x3;
  inter=t5+t3*x1+t6*x2+t7*x3;
  base=-1+t4*x1+t7*x2+t8*x3;
  y1=log(peak/base);
  y2=log(inter/base);
  fit y1 y2 / sur method=gauss converge=1.e-8 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.	27.83	0.0001	

L.M. Test: Some Preliminary Notation

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

$$\begin{aligned} D(\tilde{\theta}^*, \tilde{\Sigma}^*) \;=\; & \left\{ \frac{1}{n} \sum_{t=1}^{n} \left[\frac{\partial}{\partial \theta'} f(x_t, \tilde{\theta}^*) \right]' (\tilde{\Sigma}^*)^{-1} \left[\frac{\partial}{\partial \theta'} f(x_t, \tilde{\theta}^*) \right] \right\}^{-1} \\ & \times \left\{ \frac{1}{n} \sum_{t=1}^{n} \left[\frac{\partial}{\partial \theta'} f(x_t, \tilde{\theta}) \right]' (\tilde{\Sigma}^*)^{-1} [y_t - f(x_t, \tilde{\theta}^*)] \right\} \end{aligned}$$

$$\tilde{V}^* = \left\{ \frac{1}{n} \sum_{t=1}^{n} \left[\frac{\partial}{\partial \theta'} f(x_t, \tilde{\theta}^*) \right]' (\tilde{\Sigma}^*)^{-1} \left[\frac{\partial}{\partial \theta'} f(x_t, \tilde{\theta}^*) \right] \right\}^{-1}$$

$$\tilde{\Sigma}^* = \frac{1}{n} \sum_{t=1}^{n} \left[y_t - f(x_t, \tilde{\theta}^{\#}) \right] \left[y_t - f(x_t, \tilde{\theta}^{\#}) \right]'$$

$$\tilde{\theta}_n^* = \underset{h(\theta)=0}{\text{argmin}} \; \text{SSE}(\theta, \tilde{\Sigma}^*)$$

$$\tilde{\theta}_n^{\#} = \underset{h(\theta)=0}{\text{argmin}} \; \text{SSE}(\theta, I)$$

$$\text{SSE}(\theta, \Sigma) = \sum_{t=1}^{n} [y_t - f(x_t, \theta)]' \Sigma^{-1} [y_t - f(x_t, \theta)]$$

Note that $\tilde{\theta}_n^*$ is not $\tilde{\theta}_n$ as defined above.

Lagrange Multiplier Test

The statistic

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

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

To make degrees of freedom corrections, compare

$$R = \frac{D'(\tilde{\theta}^*, \tilde{\Sigma})(\tilde{V}^*)^{-1}D(\tilde{\theta}^*, \tilde{\Sigma}^*)}{\text{SSE}(\tilde{\theta}^*, \tilde{\Sigma}^*)/M}$$

to

$$d = \frac{nMF}{(nM-p)/q + F}$$

where F is the quantile of the F-distribution with q numerator degrees of freedom and $nM-p$ denominator degrees of freedom.

L.M. Test (SAS code)

```
proc model data=electric;
  parms t1=-2.9 t2=-1.3 t3=.82 t4=.36 t5=-1.5 t6=-1. t7=-.03 t8=-.47;
  var y1 y2 x1 x2 x3;
  peak=t1+t2*x1+t3*x2+t4*x3;
  inter=t5+t3*x1+t6*x2+t7*x3;
  base=-1+t4*x1+t7*x2+t8*x3;
  y1=log(peak/base);
  y2=log(inter/base);
  fit y1 y2 / sur method=gauss converge=1.e-8 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.           27.41     0.0001
```


Topics

- Least Squares Estimates
- Hypothesis Tests
- Confidence Intervals
- HI and HAC Variance Estimates
- Maximum Likelihood Estimators

Confidence Intervals

There is nothing special about the multivariate 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

- Least Squares Estimates
- Hypothesis Tests
- Confidence Intervals
- HI and HAC Variance Estimates
- Maximum Likelihood Estimators

Heteroskedasticity: Unknown Form

Use the multivariate nls estimator

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\text{argmin}}\ \text{SSE}(\theta, \hat{\Sigma})$$

and estimate the variance-covariance matrix of $\sqrt{n}(\hat{\theta}_n - \theta^o)$ by

$$\hat{V} = \hat{\mathcal{J}}^{-1}\hat{\mathcal{I}}\hat{\mathcal{J}}^{-1}$$

using

$$\hat{\mathcal{J}} = \frac{2}{n}\sum_{t=1}^{n}\left[\frac{\partial}{\partial\theta'}f(x_t,\hat{\theta})\right]'\hat{\Sigma}^{-1}\left[\frac{\partial}{\partial\theta'}f(x_t,\hat{\theta})\right]$$

$$\hat{\mathcal{I}}_n = \frac{4}{n}\sum_{t=1}^{n}\left[\frac{\partial}{\partial\theta'}f(x_t,\hat{\theta})\right]'\hat{\Sigma}^{-1}\hat{e}_t\hat{e}_t'\hat{\Sigma}^{-1}\left[\frac{\partial}{\partial\theta'}f(x_t,\hat{\theta})\right]$$

where

$$\hat{e} = y - f(x_t, \hat{\theta}_n)$$

Heteroskedasticity: Unknown Form, Tests

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

The likelihood ratio test cannot be used.

The Wald test is essentially $\hat{h} = h(\hat{\theta}_n)$ divided by its standard error. This can still be done:

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

where $\hat{H} = (\partial/\partial\theta')h(\hat{\theta}_n)$.

The Lagrange multiplier test is the G-N downhill direction $\tilde{D} = D(\tilde{\theta}^*, \tilde{\Sigma})$ divided by its standard error:

$$R = n\tilde{D}'\tilde{H}'(\tilde{H}\tilde{V}\tilde{H}')^{-1}\tilde{H}\tilde{D}$$

where all expressions on the previous transparency have been recomputed subject to $h(\theta) = 0$.

In both cases, reject when the statistic exceeds upper critical point of the chi squared distribution on q degrees freedom.

Serial Correlation: Unknown Form

Use the multivariate nls estimator

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\text{argmin}}\ \text{SSE}(\theta, \hat{\Sigma})$$

and estimate the variance-covariance matrix of $\sqrt{n}(\hat{\theta}_n - \theta^o)$ by

$$\hat{V} = \hat{\mathcal{J}}^{-1}\hat{\mathcal{I}}\hat{\mathcal{J}}^{-1}$$

using

$$\hat{\mathcal{J}} = \frac{2}{n}\sum_{t=1}^{n}\left[\frac{\partial}{\partial\theta'}f(x_t,\hat{\theta})\right]'\hat{\Sigma}^{-1}\left[\frac{\partial}{\partial\theta'}f(x_t,\hat{\theta})\right]$$

and $\hat{\mathcal{I}}_n$ computed as follows.

Estimation of $\mathcal{I}$

Use residuals

$$\hat{e}_t = y_t - f(x_t, \hat{\theta}_n)$$

from the multivariate nonlinear least squares estimate to compute

$$\mathcal{I}_n = \sum_{t=-l(n)}^{l(n)} w\left(\frac{\tau}{l(n)}\right) \hat{\mathcal{I}}_{n\tau}$$

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

$$\hat{\mathcal{I}}_{n\tau} = \begin{cases} \frac{4}{n} \sum_{t=\tau+1}^{n} \left[\frac{\partial}{\partial\theta'} f(x_t, \hat{\theta})\right]' \hat{\Sigma}^{-1} \hat{e}_t \hat{e}'_{t-\tau} \hat{\Sigma}^{-1} \left[\frac{\partial}{\partial\theta'} f(x_{t-\tau}, \hat{\theta})\right] & \tau \geq 0 \\ \hat{\mathcal{I}}'_{n,-\tau} & \tau < 0 \end{cases}$$

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

Serial Correlation: Unknown Form, Tests

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

The likelihood ratio test cannot be used.

The Wald test is essentially $\hat{h} = h(\hat{\theta}_n)$ divided by its standard error. This can still be done:

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

where $\hat{H} = (\partial/\partial\theta')h(\hat{\theta}_n)$.

The Lagrange multiplier test is the G-N downhill direction $\tilde{D} = D(\tilde{\theta}^*, \tilde{\Sigma})$ divided by its standard error:

$$R = n\tilde{D}'\tilde{H}'(\tilde{H}\tilde{V}\tilde{H}')^{-1}\tilde{H}\tilde{D}$$

where all expressions on the previous transparencies have been recomputed subject to $h(\theta) = 0$.

In both cases, reject when the statistic exceeds upper critical point of the chi squared distribution on q degrees freedom.

Topics

- Least Squares Estimates
- Hypothesis Tests
- Confidence Intervals
- HI and HAC Variance Estimates
- Maximum Likelihood Estimators

Maximum Likelihood: Statistical Model

$$y_t = f(x_t, \theta) + e_t \quad t = 1, 2, \ldots, n$$

- y_t the dependent variable, M-variate, observed
- x_t the explanatory variables, k-variate, observed
- θ model parameters, p-variate, unknown (to be estimated)
- e_t the errors are iid $N_M(0, \Sigma)$

As previously, let

$$\text{SSE}(\theta, \Sigma) = \sum_{t=1}^{n} [y_t - f(x_t, \theta)]' \Sigma^{-1} [y_t - f(x_t, \theta)]$$

Maximum Likelihood Estimator: Computation

If

$$\begin{aligned}
\hat\theta_0 &= \underset{\theta\in\Theta}{\text{argmin}}\ \text{SSE}(\theta, I)\\
\hat\Sigma_0 &= \frac{1}{n}\sum_{t=1}^{n}[y_t - f(x_t,\hat\theta_0)][y_t - f(x_t,\hat\theta_0)]'\\
\hat\theta_1 &= \underset{\theta\in\Theta}{\text{argmin}}\ \text{SSE}(\theta, \hat\Sigma_0)\\
\hat\Sigma_1 &= \frac{1}{n}\sum_{t=1}^{n}[y_t - f(x_t,\hat\theta_1)][y_t - f(x_t,\hat\theta_1)]'\\
\hat\theta_2 &= \underset{\theta\in\Theta}{\text{argmin}}\ \text{SSE}(\theta, \hat\Sigma_1)\\
&\vdots
\end{aligned}$$

then

$$\hat\theta_\infty = \lim_{n\to\infty}\hat\theta_i \quad \hat\Sigma_\infty = \lim_{n\to\infty}\hat\Sigma_i$$

exist and are the maximum likelihood estimators of θ^o and Σ^o. Also, $\text{SSE}(\hat\theta_\infty, \hat\Sigma_\infty) = M$.

Maximum Likelihood Estimation

While the above ziq-zag algorithm is not the best algorithm for computing the maximum likelihood estimator, it does show that the maximum likelihood estimator is also a generalized least squares estimator and therefore all our previous iid results apply with $\hat\theta_\infty$ and $\hat\Sigma_\infty$ replacing $\hat\theta$ and $\hat\Sigma$.

Therefore, the tests for H : $h(\theta^o) = 0$ against A : $h(\theta^o) \neq 0$ discussed above for the iid case are applicable.

Maximum likelihood estimation also provides the means to test hypotheses about the variance H : $h(\Sigma^o) = 0$ or even joint hypotheses H : $h(\theta^o, \Sigma^o) = 0$. These are not much used, so we will skip them in lecture. They are discussed in NLSM.

Maximum Likelihood Estimation: Caveat

Similarly to the definition of $\hat\theta_\infty$ and $\hat\Sigma_\infty$, let

$$\begin{aligned}
\tilde\theta_0 &= \underset{h(\theta)=0}{\text{argmin}}\ \text{SSE}(\theta, I)\\
\tilde\theta_i &= \underset{h(\theta)=0}{\text{argmin}}\ \text{SSE}(\theta, \tilde\Sigma_{i-1})\\
\tilde\Sigma_i &= \frac{1}{n}\sum_{t=1}^{n}[y_t - f(x_t,\tilde\theta_i)][y_t - f(x_t,\tilde\theta_i)]'
\end{aligned}$$

$$\tilde\theta_\infty = \lim_{n\to\infty}\tilde\theta_i \quad \tilde\Sigma_\infty = \lim_{n\to\infty}\tilde\Sigma_i$$

The likelihood ratio test statistic for

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

is

$$L = n(\log\det\tilde\Sigma_\infty - \log\det\hat\Sigma_\infty)$$

not the "likelihood ratio" test statistic

$$L = S(\tilde\theta_\infty, \hat\Sigma_\infty) - S(\hat\theta_\infty, \hat\Sigma_\infty)$$

used in the generalized least squares theory. Although they are actually the same to within a first order Taylor's expansion, it is better to use the former with mle estimates to avoid confusing people.

Topics

- Least Squares Estimates
- Hypothesis Tests
- Confidence Intervals
- HI and HAC Variance Estimates
- Maximum Likelihood Estimators