NONLINEAR STATISTICAL MODELS

by

A. Ronald Gallant

CHAPTER 6. Multivariate Nonlinear Regression

NOT FOR GENERAL DISTRIBUTION

This copy is reproduced by special permission of the author and publisher from:

A. Ronald Gallant. Nonlinear Statistical Models. New York: John Wiley and Sons, Inc, forthcoming.

and is being circulated privately for discussion purposes only. Reproduction from this copy is expressly prohibited.

Please send comments and report errors to the following address:

A. Ronald Gallant
Institute of Statistics
North Carolina State University
Post Office Box 8203
Raleigh, NC 27695-8203

Phone 1-919-737-2531

NONLINEAR STATISTICAL MODELS

Table of Contents

Table of Contents (continued)

Chapter 6. Multivariate Nonlinear Regression

All that separates multivariate regression from univariate regression is a linear transformation. Accordingly, the main thrust of this chapter is to identify the transformation, to estimate it, and then to apply the ideas of Chapter 1. In Chapter 1 we saw that there is little difference between linear and nonlinear least squares save for some extra tedium in the computations. We saw that if one uses the likelihood ratio test to test hypothesis and construct confidence intervals then inferences are reliable provided one uses the same degrees of freedom corrections used in linear regression and provided that the hypothesis of spherically distributed errors is reasonably accurate. These are the main ideas of this chapter.

1. INTRODUCTION

In Chapter 1 we considered univariate nonlinear model

$$y_t = f(x_t, \theta^\circ) + e_t \qquad t = 1, 2, \ldots, n .$$

Here we consider the case where there are M such regressions

$$y_{\alpha t} = f_\alpha(x_t, \theta_\alpha^\circ) + e_{\alpha t} \qquad t = 1, 2, \ldots, n ; \ \alpha = 1, 2, \ldots, M$$

that are related in one of two ways. The first arises most naturally when

$$y_{\alpha t} \qquad \alpha = 1, 2, \ldots, M$$

represent repeated measures on the same subject, height and weight measurements on the same individual for instance. In this case one would expect the observations with the same t index to be correlated, viz

$$C(y_{\alpha t}, y_{\beta t}) = \sigma_{\alpha\beta} \quad .$$

One often refers to this situation as contemporaneous correlation. The second way these regressions can be related is through shared parameters. Stacking the parameter vectors and writing

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_M \end{pmatrix}$$

one can have

$$\theta = g(\rho)$$

where ρ has smaller dimension than θ . If either or both of these relationships obtain, contemporaneous correlation or shared parameters, estimators with improved efficiency can be obtained; improved in the sense of better efficiency than that which obtains by applying the methods of Chapter 1 M times (Problem 12, Section 3). An example that exhibits these characteristics that we shall use heavily for illustration is the following.

EXAMPLE 1. (Consumer Demand) The data shown in Tables 1a and 1b is to be transformed as follows

$$y_1 = \ell n \text{ (peak expenditure share)} - \ell n \text{ (base expenditure share)}$$
$$y_2 = \ell n \text{ (intermediate expenditure share)} - \ell n \text{ (base expenditure share)}$$
$$x_1 = \ell n \text{ (peak price/expenditure)}$$
$$x_2 = \ell n \text{ (intermediate price/expenditure)}$$
$$x_2 = \ell n \text{ (base price/expenditure)} .$$

As notation, set

$$y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \qquad x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

$$y_t = \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} \qquad x_t = \begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} \qquad t = 1, 2, \ldots, 224$$

These data are presumed to follow the model

$$y_{1t} = \ell n[(a_1 + x'b_{(1)})/(a_3 + x'b_{(3)})] + e_{1t}$$
$$y_{2t} = \ell n[(a_2 + x'b_{(2)})/(a_3 + x'b_{(3)})] + e_{2t}$$

where

$$a = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} ,$$

$$B = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}$$

and $b'_{(i)}$ denotes the ith row of B, viz.

$$b'_{(i)} = (b_{i1}, b_{i2}, b_{i3}) \ .$$

The errors

$$e_t = \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}$$

are assumed to be independently and identically distributed each with mean zero and variance-covariance matrix Σ .

There are various hypotheses that one might impose on the model. Two are of the nature of maintained hypotheses that follow directly from the theory of demand and ought to be satisfied. These are:

H_1: a_3 and $b_{(3)}$ are the same in both equations, as the notation suggests.

H_2: B is a symmetric matrix.

There is a third hypothesis that would be a considerable convenience if it were true

$$H_3 \colon \Sigma_{i=1}^{3} a_i = -1, \ \Sigma_{j=1}^{3} b_{ij} = 0 \quad \text{for } i = 1, 2, 3 \ .$$

The theory supporting this model specification follows; the reader who has no interest in the theory can skip over the rest of the example.

The theory of consumer demand is fairly straightforward. Given an income Y which can be spent on N different goods which sell at prices $p_1, p_2, \ldots, p_N$ the consumer's problem is to decide what quantities $q_1, q_2, \ldots, q_N$ of each good to purchase. One assumes that the consumer has the ability to rank various bundles of goods in order of preference. Denoting a bundle by the vector

$$q = (q_1, q_2, \ldots, q_N) ,$$

the assumption of an ability to rank bundles is equivalent to the assumption that there is a (utility) function $u(q)$ such that $u(q^\circ) > u(q^*)$ means bundle q° is preferred to bundle q^*. Since a bundle costs $p'q$ with $p' = (p_1, p_2, \ldots, p_N)$ the consumers problem is

$$\text{maximize } u(q)$$
$$\text{subject to } p'q = Y .$$

This is the same problem as

$$\text{maximize } u(q)$$
$$\text{subject to } (p/Y)'q = 1$$

which means that the solution must be of the form

$$q = q(v)$$

with $v = p/Y$. The function $q(v)$ mapping the positive orthant of R^N into the positive orthant of R^N is called the consumer's demand system. It is usually assumed in applied work that all prices are positive and that a bundle with some $q_i = 0$ is never chosen.

If one substitutes the demand system $q(v)$ back into the utility function one obtains the function

$$g(v) = u[q(v)]$$

which gives the maximum utility that a consumer can achieve at the price/income point v. The function $g(v)$ is called the indirect utility function. A property of the indirect utility function that makes it extremely useful in applied work is that the demand system is proportional to the gradient of the indirect utility function (Deaton and Muellbauer, 1980), viz.

$$q(v) = (\partial/\partial v)g(v)/v'(\partial/\partial v)g(v) \,.$$

This relationship is called Roy's identity. Thus, to implement the theory of consumer demand one need only specify a parametric form $g(v|\theta)$ and then fit the system

$$q = (\partial/\partial v)g(v|\theta)/v'(\partial/\partial v)g(v|\theta)$$

to observed values of (q,v) in order to estimate θ . The theory asserts that the fitted function $g(v|\theta)$ should be decreasing in each argument, $(\partial/\partial v_i)g(v|\theta) < 0$, and should be quasi-convex, $v'(\partial^2/\partial v \partial v')g(v|\theta)v > 0$ for every v with $v'(\partial/\partial v)g(v|\theta) = 0$ (Deaton and Muellbauer, 1980). If $g(v|\theta)$ has this property then there exists a corresponding $u(q)$. Thus, in applied work, there is no need to bother with $u(q)$; $g(v|\theta)$ is enough.

It is easier to arrive at a stochastic model if we reexpress the demand system in terms of expenditure shares. Accordingly let $\mathrm{diag}(v)$ denote a diagonal matrix with the components of the vector v along the diagonal and set

$$s = \mathrm{diag}(v)q$$

$$s(v|\theta) = \mathrm{diag}(v)\ (\partial/\partial v)g(v|\theta)/v'(\partial/\partial v)g(v|\theta) \,.$$

Observe that

$$s_i = v_i q_i = p_i q_i / Y$$

so that s_i denotes that proportion of total expenditure Y spent on the ith good. As such $1's = \Sigma_{i=1}^{N} s_i = 1$ and $1's(v|\theta) = 1$.

The deterministic model suggests that the distribution of the shares has a location parameter that depends on $s(v|\theta)$ in a simple way. What seems to the case with this sort (Rossi, 1983) of data is that observed shares follow the logistic-normal distribution (Aitchison and Shen, 1980) with location parameter

$$\mu = \ell n\ s(v|\theta)$$

where $\ell n\ s(v|\theta)$ denotes the N-vector with components $\ell n\ s_i(v|\theta)$ for $i = 1, 2, \ldots, N$. The logistic-normal distribution is characterized as follows. Let w be normally distributed with mean vector μ and a variance-covariance matrix $C(w,w')$ that satisfies $1'C(w,w')1 = 0$. Then s has the logistic-normal distribution if

$$s = e^{w}/(\Sigma_{i=1}^{N}\ e^{w_i})$$

where e^w denotes the vector with components e^{w_i} for $i = 1, 2, \ldots, N$. A log transform yields

$$\ell n\ s = w - \ell n(\Sigma_{i=1}^{N} e^{w_i})1$$

whence

$$\ell n\ s_i - \ell n\ s_N = w_i - w_N\ , \qquad i = 1, 2, \ldots, N\text{-}1\ .$$

Writing $w_i - w_N = \mu_i - \mu_N + e_i$ for $i = 1, 2, \ldots, N\text{-}1$ we have equations that can be fit to data

$$\ell n\ s_i - \ell n\ s_N = \ell n\ (\partial/\partial v_i)g(v|\theta) \quad \ell n\ (\partial/\partial v_N)g(v|\theta) + e_i \qquad i = 1,2,\ldots,N\text{-}1\ .$$

The last step in implementing this model is the specification of a functional form for $g(v|\theta)$. Theory implies a strong preference for a low order multivariate Fourier series expansion (Gallant, 1981, 1982; Elbadawi, Gallant, and Souza, 1983) but since our purpose is illustrative the choice will be governed by simplicity and manipulative convenience. Accordingly, let $g(v|\theta)$ be specified as the Translog (Christensen, Jorgenson, and Lau, 1975)

$$g(v|\theta) = \Sigma_{i=1}^{N} a_i \ell n(v_i) + \tfrac{1}{2}\Sigma_{i=1}^{N}\Sigma_{j=1}^{N} b_{ij}\ \ell n(v_i)\ \ell n(v_j)$$

or

$$g(v|\theta) = a'x + (\tfrac{1}{2})x'B\,x$$

with $x = \ell n\ v$ and

$$a' = (a_1,\ a_2,\ a_3)$$

$$B = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} \quad .$$

Differentiation yields

$$(\partial/\partial v)g(v|\theta) = [\mathrm{diag}(v)]^{-1}[a + \tfrac{1}{2}\ (B + B')x]\ .$$

One can see from this expression that B can be taken to be symmetric without loss of generality. With this assumption we have

$$(\partial/\partial v)g(v|\theta) = [\mathrm{diag}(v)]^{-1}(a + Bx)\ .$$

Recall that in general shares are computed as

$$s(v|\theta) = \mathrm{diag}(v)\ (\partial/\partial v)g(v|\theta)/v'(\partial/\partial v)g(v|\theta)$$

which reduces to

$$s(v|\theta) = (a + Bx)/1'(a + Bx)$$

in this instance. Differenced log shares are

$$\ell n\ s_i(v|\theta) - \ell n\ s_N(v|\theta) = \ell n[(a_i + x'b_{(i)})/(a_N + x'b_{(N)})]\ .$$

The model set forth in the beginning paragraphs of this discussion follows from the above equation. The origins of hypotheses H_1 and H_2 are apparent as well.

One notes, however, that we applied this model not to all goods $q_1, q_2, \ldots, q_N$ and income Y but rather to three categories of electricity expenditure - peak = q_1, intermediate = q_2, base = q_3 - and to total electricity expenditure E. A (necessary and sufficient) condition that permits one to apply the theory of demand essentially intact to the electricity subsystem, as we have done, is that the utility function is of the form (Blacorby, Primont, and Russell, 1978, Ch. 5)

$$u[u_{(1)}(q_1, q_2, q_3), q_4, \ldots, q_N] .$$

If the utility function is of this form and E is known it is fairly easy to see that optimal allocation of E to q_1, q_2, and q_3 can be computed by solving

$$\begin{aligned}&\text{maximize } u_{(1)}(q_1, q_2, q_3)\\ &\text{subject to } \Sigma_{i=1}^{3} p_i q_i = E .\end{aligned}$$

Since this problem has exactly the same structure as the original problem, one just applies the previous theory with $N = 3$ and $Y = E$.

There is a problem in passing from the deterministic version of the subsystem to the stochastic specification. One usually prefers to regard prices and income as independent variables and condition the analysis on p and Y . Expenditure in the subsystem, from this point of view, is to be regarded as stochastic with a location parameter depending on p, Y and possibly on demographic characteristics, viz

$$E = f(p, Y, \text{etc.}) + \text{error} .$$

For now, we shall ignore this problem, implicitly treating it as an errors in variables problem of negligible consequence. That is, we assume that in observing E we are actually observing f(p, Y, etc.) with negligible error so that an analysis conditioned on E will be adequate. In Chapter 8 we shall present methods that take formal account of this problem.

In this connection, hypothesis H_3 implies that $g(v|\theta)$ is homogeneous of degree one in v which in turn implies that the first-stage allocation function has the form

$$f(p, Y, \text{etc.}) = f[\Pi(p_1, p_2, p_3), p_4, \ldots, p_N, Y, \text{etc.}]$$

where $\Pi(p_1, p_2, p_3)$ is a price index for electricity which must, itself, be homogeneous of degree one in p_1, p_2, p_3 (Blackorby, Primont, and Russell, 1978, Ch. 5). This leads to major simplifications in the interpretation of results which see Caves and Christensen (1980).

One word of warning regarding Table 1c, all data is constructed following the protocol described in Gallant and Koenker (1984) save income. Some income values have been imputed by prediction from a regression equation. These values can be identified as those not equal to one of the values 500, 1500, 2500, 3500, 4500, 5500, 7000, 9000, 11000, 13500, 17000, 22500, 27500, 40000, 70711. The listed values are the means of the questionnaire's class boundaries save the last which is the mean of an open ended interval assuming that income follows the Pareto distribution. The prediction equation includes variables not shown in Table 1c, namely age and years of education of a member of the household, the respondent or head in most instances. □

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

t	Treatment	Expenditure Share: Base	Intermediate	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
23	2	0.165714	0.276980	0.557305	0.84658
24	2	0.145370	0.173112	0.681518	1.60788
25	2	0.184467	0.268865	0.546668	0.73838
26	2	0.162269	0.280939	0.556792	0.81116
27	2	0.112016	0.220850	0.667133	2.01503
28	2	0.226863	0.257833	0.515304	2.32035
29	2	0.118028	0.219830	0.662142	2.40172
30	2	0.137761	0.345117	0.517122	0.57141
31	2	0.079115	0.257319	0.663566	0.94474
32	2	0.185022	0.265051	0.549928	1.63778
33	2	0.144524	0.276133	0.579343	0.75816
34	2	0.201734	0.241966	0.556300	1.00136
35	2	0.094890	0.227651	0.677459	1.11384
36	2	0.102843	0.264515	0.632642	1.07185
37	2	0.107760	0.214232	0.678009	1.53659
38	2	0.156552	0.236422	0.607026	0.24099
39	2	0.088431	0.222746	0.688822	0.58066
40	2	0.146236	0.301884	0.551880	2.52983
41	3	0.080802	0.199005	0.720192	1.14741
42	3	0.100711	0.387758	0.511531	0.97934
43	3	0.073483	0.335280	0.591237	1.09361
44	3	0.059455	0.259823	0.680722	2.19468
45	3	0.076195	0.378371	0.545434	1.98221

(Continued next page)

Table 1a. (Continued).

t	Treatment	Expenditure Share Base	Expenditure Share Intermediate	Expenditure Share Peak	Expenditure ($ per day)
46	3	0.076926	0.325032	0.598042	1.78194
47	3	0.086052	0.339653	0.574295	3.24274
48	3	0.069359	0.278369	0.652272	0.47593
49	3	0.071265	0.273866	0.654869	1.38369
50	3	0.100562	0.306247	0.593191	1.57831
51	3	0.050203	0.294285	0.655513	2.16900
52	3	0.059627	0.311932	0.628442	2.11575
53	3	0.081433	0.328604	0.589962	0.35681
54	3	0.075762	0.285972	0.638265	1.55275
55	3	0.042910	0.372337	0.584754	1.06305
56	3	0.086846	0.340184	0.572970	4.02013
57	3	0.102537	0.335535	0.561928	0.60712
58	3	0.068766	0.310782	0.620452	1.15334
59	3	0.058405	0.307111	0.634485	2.43797
60	4	0.055227	0.300839	0.643934	0.10082
61	4	0.107435	0.273937	0.618628	0.69302
62	4	0.105958	0.291205	0.602837	1.12592
63	4	0.132278	0.279429	0.588293	1.84425
64	4	0.094195	0.328866	0.576940	1.57972
65	4	0.115259	0.401079	0.483663	1.27034
66	4	0.150229	0.317866	0.531905	0.56330
67	4	0.168780	0.307669	0.523551	3.43139
68	4	0.118222	0.318080	0.563698	1.00979
69	4	0.103394	0.307671	0.588936	2.08458
70	4	0.124007	0.362115	0.513879	1.30410
71	4	0.197987	0.280130	0.521884	3.48146
72	4	0.108083	0.337004	0.554913	0.53206
73	5	0.088798	0.232568	0.678634	3.28987
74	5	0.100508	0.272139	0.627353	0.32678
75	5	0.127303	0.298519	0.574178	0.52452
76	5	0.109718	0.228172	0.662109	0.36622
77	5	0.130080	0.231037	0.638883	0.63788
78	5	0.148562	0.323579	0.527859	1.42239
79	5	0.106306	0.252137	0.641556	0.93535
80	5	0.080877	0.214172	0.704951	1.26243
81	5	0.081810	0.135665	0.782525	1.51472
82	5	0.131749	0.278338	0.589913	2.07858
83	5	0.059180	0.254533	0.686287	1.60681
84	5	0.078620	0.267252	0.654128	1.54706
85	5	0.090220	0.293831	0.615949	2.61162
86	5	0.086916	0.193967	0.719117	2.96418
87	5	0.132383	0.230489	0.637127	0.26912
88	5	0.085560	0.252321	0.662120	0.42554
89	5	0.071368	0.276238	0.652393	1.01926
90	5	0.061196	0.245025	0.693780	1.53807

(Continued next page)

Table 1a. (Continued).

t	Treatment	Expenditure Share: Base	Expenditure Share: Intermediate	Expenditure Share: Peak	Expenditure ($ per day)
91	5	0.086608	0.233981	0.679411	0.75711
92	5	0.105628	0.305471	0.588901	0.83647
93	5	0.078158	0.202536	0.719307	1.92096
94	5	0.048632	0.216807	0.734560	1.57795
95	5	0.094527	0.224344	0.681128	0.83216
96	5	0.092809	0.209154	0.698037	1.39364
97	5	0.035751	0.166231	0.798018	1.72697
98	5	0.065205	0.205058	0.729736	2.04120
99	5	0.092561	0.193848	0.713591	2.04708
100	5	0.063119	0.234114	0.702767	3.43969
101	5	0.091186	0.224488	0.684326	2.66918
102	5	0.047291	0.262623	0.690086	2.71072
103	5	0.081575	0.206400	0.712025	3.36803
104	5	0.108165	0.243650	0.648185	0.65682
105	5	0.079534	0.320450	0.600017	0.95523
106	5	0.084828	0.247189	0.667984	0.61441
107	5	0.063747	0.210343	0.725910	1.85034
108	5	0.081108	0.249960	0.668932	2.11274
109	5	0.089942	0.206601	0.703457	1.54120
110	5	0.046717	0.224784	0.728499	3.54351
111	5	0.114925	0.272279	0.612796	2.61769
112	5	0.115055	0.264415	0.620530	3.00236
113	5	0.081511	0.223870	0.694618	1.74166
114	5	0.109658	0.343593	0.546750	1.17640
115	5	0.114263	0.304761	0.580976	0.74566
116	5	0.115089	0.226412	0.658499	1.30392
117	5	0.040622	0.198986	0.760392	2.13339
118	5	0.073245	0.238522	0.688234	2.83039
119	5	0.087954	0.287450	0.624596	1.62179
120	5	0.091967	0.206131	0.701902	2.18534
121	5	0.142746	0.302939	0.554315	0.26503
122	5	0.117972	0.253811	0.628217	0.05082
123	5	0.071573	0.248324	0.680103	0.42740
124	5	0.073628	0.290586	0.635786	0.47979
125	5	0.121075	0.350781	0.528145	0.59551
126	5	0.077335	0.339358	0.583307	0.47506
127	5	0.074766	0.167202	0.758032	2.11867
128	5	0.208580	0.331363	0.460058	1.13621
129	5	0.080195	0.210619	0.709185	2.61204
130	5	0.066156	0.204118	0.729726	1.45227
131	5	0.112282	0.252638	0.635080	0.79071
132	5	0.041310	0.093106	0.865584	1.30697
133	5	0.102675	0.297009	0.600316	0.93691
134	5	0.102902	0.270832	0.626266	0.98718
135	5	0.118932	0.250104	0.630964	1.40085

(Continued next page)

Table 1a. (Continued).

t	Treatment	Expenditure Share: Base	Expenditure Share: Intermediate	Expenditure Share: Peak	Expenditure ($ per day)
136	5	0.139760	0.322394	0.537846	1.78710
137	5	0.121616	0.214626	0.663758	8.46237
138	5	0.065701	0.263818	0.670481	1.58663
139	5	0.034029	0.175181	0.790790	2.62535
140	5	0.074476	0.194744	0.730780	4.29430
141	5	0.059568	0.229705	0.710727	0.65404
142	5	0.088128	0.295546	0.616326	0.41292
143	5	0.075522	0.213622	0.710856	2.02370
144	5	0.057089	0.195720	0.747190	1.76998
145	5	0.096331	0.301692	0.601977	0.99891
146	5	0.120824	0.250280	0.628896	0.27942
147	6	0.034529	0.193456	0.772015	0.91673
148	6	0.026971	0.180848	0.792181	1.15617
149	6	0.045271	0.141894	0.812835	1.57107
150	6	0.067708	0.219302	0.712990	1.24515
151	6	0.079335	0.230693	0.689972	1.70748
152	6	0.022703	0.178896	0.798401	1.79959
153	6	0.043053	0.157142	0.799805	4.61665
154	6	0.057157	0.245931	0.696912	0.59504
155	6	0.063229	0.136192	0.800579	1.42499
156	6	0.076873	0.214209	0.708918	1.34371
157	6	0.027353	0.124894	0.847753	2.74908
158	6	0.067823	0.146994	0.785183	1.84628
159	6	0.056388	0.189185	0.754428	3.82472
160	6	0.036841	0.194994	0.768165	1.18199
161	6	0.059160	0.138681	0.802158	2.07338
162	6	0.051980	0.215700	0.732320	0.80376
163	6	0.027300	0.145072	0.827628	1.52316
164	6	0.014790	0.179619	0.805591	3.17526
165	6	0.047865	0.167561	0.784574	3.30794
166	6	0.115629	0.231381	0.652990	0.72456
167	7	0.104970	0.147525	0.747505	0.50274
168	7	0.119254	0.187409	0.693337	1.22571
169	7	0.042564	0.112839	0.844596	2.13534
170	7	0.096756	0.150178	0.753066	5.56011
171	7	0.063013	0.168422	0.768565	3.11725
172	7	0.080060	0.143934	0.776006	0.99796
173	7	0.097493	0.173391	0.729116	0.67859
174	7	0.102526	0.220954	0.676520	0.79027
175	7	0.085538	0.195686	0.718776	2.24498
176	7	0.068733	0.166248	0.765019	2.01993
177	7	0.094915	0.140119	0.764966	4.07330
178	7	0.076163	0.132046	0.791792	3.66432
179	7	0.099943	0.176885	0.723172	0.40768
180	7	0.081494	0.175082	0.743425	1.09065

(Continued next page)

Table 1a. (Continued).

t	Treatment	Expenditure Share: Base	Intermediate	Peak	Expenditure ($ per day)
181	7	0.196026	0.299348	0.504626	1.35008
182	7	0.093173	0.235816	0.671011	1.06138
183	7	0.172293	0.173032	0.654675	0.99219
184	7	0.067736	0.159600	0.772663	3.69199
185	7	0.102033	0.171697	0.726271	2.36676
186	7	0.067977	0.151109	0.780914	1.84563
187	8	0.071073	0.238985	0.689942	0.18316
188	8	0.049453	0.286788	0.663759	2.23986
189	8	0.062748	0.255129	0.682123	3.48084
190	8	0.032376	0.154905	0.812719	7.26135
191	8	0.055055	0.225296	0.719648	1.68814
192	8	0.037829	0.179051	0.783120	1.13804
193	8	0.020102	0.172396	0.807502	1.40894
194	8	0.021917	0.149092	0.828992	3.47472
195	8	0.047590	0.174735	0.777675	3.37689
196	8	0.063446	0.235823	0.700731	3.14810
197	8	0.034719	0.159398	0.805883	3.21710
198	8	0.055428	0.200488	0.744084	1.13941
199	8	0.058074	0.254823	0.687103	2.55414
200	8	0.060719	0.209763	0.729518	0.29071
201	8	0.045681	0.206177	0.748142	1.21336
202	8	0.040151	0.263161	0.696688	1.02370
203	8	0.072230	0.281460	0.646310	1.40580
204	8	0.064366	0.269816	0.665819	0.97704
205	8	0.035993	0.191422	0.772585	2.09909
206	9	0.091638	0.215290	0.693073	1.03679
207	9	0.072171	0.236658	0.691171	2.36788
208	9	0.056187	0.195345	0.748468	3.45908
209	9	0.095888	0.229586	0.674526	3.63796
210	9	0.069809	0.219558	0.710633	2.56887
211	9	0.142920	0.223801	0.633279	2.00319
212	9	0.087323	0.196401	0.716276	2.40644
213	9	0.064517	0.218711	0.716772	2.58552
214	9	0.086882	0.194778	0.718341	8.94023
215	9	0.067463	0.219228	0.713309	3.75275
216	9	0.105610	0.230661	0.663730	0.34082
217	9	0.138992	0.283123	0.577885	1.62649
218	9	0.081364	0.186967	0.731670	2.31678
219	9	0.114535	0.221751	0.663714	1.77709
220	9	0.069940	0.280622	0.649438	1.38765
221	9	0.073137	0.143219	0.783643	3.46442
222	9	0.096326	0.243241	0.660434	1.74696
223	9	0.083284	0.202951	0.713765	1.28613
224	9	0.179133	0.299403	0.521465	1.15897

Source: Gallant and Koenker(1984)

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

Treatment	Price (cents per kwh)		
	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.

Table 1c. Consumer Demographic Characteristics.

t	Family Size	Income ($ per yr)	Residence Size (SqFt)	Residence Heat Loss (Btuh)	Elec. Range (1=yes)	Washer (1=yes)	Dryer (1=yes)	Air Condition. Central (1=yes)	Air Condition. Window (Btuh)
1	2	17000	600	4305	0	1	0	0	13000
2	6	13500	900	9931	1	1	0	0	0
3	2	7000	1248	18878	1	1	0	0	0
4	3	11000	1787	17377	1	1	0	0	0
5	4	27500	2900	24894	1	0	0	1	5000
6	3	13500	2000	22526	1	1	1	0	24000
7	4	22500	3800	17335	1	1	1	1	0
8	7	3060	216	4496	1	0	0	0	0
9	3	7000	1000	8792	0	1	1	0	18000
10	1	6793	1200	14663	0	0	0	0	.
11	5	11000	1000	14480	1	1	0	0	0
12	5	17000	704	3192	1	1	1	1	24000
13	3	5500	2100	8631	1	1	0	1	0
14	2	13500	1400	19720	1	1	1	0	19000
15	4	22500	1252	7386	1	1	1	0	24000
16	7	17000	916	7194	0	1	0	0	0
17	2	11000	1800	17957	1	1	1	1	0
18	2	13500	780	4641	1	1	0	1	0
19	3	6570	960	11396	1	1	0	0	24000
20	4	9000	768	8195	1	1	1	0	0
21	2	11000	1200	7812	1	1	1	1	10000
22	4	13500	900	8878	1	1	1	1	0
23	3	40000	2200	15098	1	1	1	0	0
24	5	7000	1000	7041	1	1	0	0	10000
25	3	13500	720	5130	0	1	1	0	0
26	2	13500	550	7532	1	1	0	0	12000
27	4	17000	1600	9674	1	1	1	1	0
28	4	27500	2300	13706	1	1	0	1	0
29	6	15797	1000	10372	1	1	1	0	10000
30	2	11000	880	7477	0	1	1	0	19000
31	4	9000	1200	14013	1	1	1	0	0
32	4	17052	2200	15230	1	1	0	0	0
33	2	14812	1080	13170	1	0	0	0	0
34	3	27500	870	10843	1	1	1	0	18500
35	2	4562	800	9393	1	1	1	0	6000
36	2	7000	1200	11395	1	1	0	0	0
37	3	9000	900	6175	1	1	0	0	23000
38	2	4711	1500	17655	1	0	0	0	0
39	5	14652	1500	11916	1	1	1	0	0
40	4	70711	2152	16552	1	1	1	1	0
41	2	7000	832	4316	1	1	1	1	0
42	3	22500	1700	9209	1	1	1	1	0
43	11	4500	1248	9607	1	1	0	0	0
44	5	11000	1808	19400	1	1	1	0	28000
45	6	22500	1800	19981	1	1	1	1	0

(Continued next page).

Table 1c. (Continued).

t	Family Size	Income ($ per yr)	Residence Size (SqFt)	Residence Heat Loss (Btuh)	Elec. Range (1=yes)	Washer (1=yes)	Dryer (1=yes)	Air Condition. Central (1=yes)	Air Condition. Window (Btuh)
46	4	22500	1800	18573	0	0	0	1	0
47	3	40000	4200	16264	1	1	1	1	0
48	2	9000	1400	10541	1	1	1	0	24000
49	2	13500	2500	29231	1	1	0	0	16000
50	6	17000	1300	5805	1	1	1	0	21000
51	3	11000	780	5894	1	1	1	1	0
52	1	4500	1000	13714	0	0	0	0	6000
53	2	11267	960	7863	1	1	0	0	0
54	3	2500	1000	12973	1	1	0	0	0
55	1	7430	1170	9361	1	1	1	0	0
56	4	17000	2900	12203	1	1	1	1	0
57	1	22500	1000	10131	0	1	0	0	0
58	3	22500	1250	12773	1	1	1	0	12000
59	3	7000	1400	11011	1	1	1	0	29000
60	1	2500	835	12730	1	0	0	0	0
61	1	13500	1300	7196	1	1	0	0	32000
62	7	11000	540	7798	1	1	0	0	0
63	4	14381	1100	8700	1	1	1	0	30000
64	2	9000	900	5726	1	0	0	0	12000
65	3	11000	720	3854	1	1	1	1	0
66	5	5500	780	6236	1	1	0	1	0
67	4	40000	1450	8160	1	1	1	0	28000
68	2	3500	1100	10102	1	1	0	0	12000
69	2	17000	3000	36124	1	1	0	1	0
70	4	11000	1534	15711	1	0	0	0	0
71	2	40000	2000	11250	1	1	1	1	0
72	2	2500	1400	15040	0	0	0	0	6000
73	4	17000	1400	13544	1	0	1	1	0
74	2	1500	656	7383	1	0	0	0	0
75	3	9000	772	13229	1	0	0	0	1800
76	1	9000	600	4035	1	1	0	0	0
77	5	5500	500	6110	1	0	0	0	0
78	3	13500	1200	11097	1	1	1	0	10000
79	2	13590	1300	12869	1	0	0	0	24000
80	4	11000	1045	11224	1	1	0	0	0
81	2	9687	768	7565	1	1	1	0	10000
82	2	17000	1100	9159	0	1	1	0	10000
83	11	4500	480	6099	1	1	0	0	0
84	5	13500	1976	12498	1	1	1	0	0
85	4	40000	2500	23213	1	1	1	0	.
86	5	22500	2100	12314	1	1	1	1	0
87	3	3500	1196	14725	0	0	0	0	0
88	3	12100	950	11174	0	0	0	0	0
89	3	3500	1080	12186	1	0	0	0	0
90	2	7000	1400	10050	1	1	0	0	28000

(Continued next page).

Table 1c. (Continued).

t	Family Size	Income ($ per yr)	Residence Size (SqFt)	Residence Heat Loss (Btuh)	Elec. Range (1=yes)	Washer (1=yes)	Dryer (1=yes)	Air Condition. Central (1=yes)	Air Condition. Window (Btuh)
91	2	3500	1800	16493	1	1	1	0	2000
92	2	7000	1456	17469	0	1	0	0	18000
93	4	9000	1100	6177	1	1	1	0	23000
94	2	3500	1500	21659	1	1	1	0	18000
95	4	9894	720	6133	1	1	1	0	6000
96	1	22500	1500	7952	1	0	0	1	0
97	4	13500	1500	10759	1	0	1	1	0
98	4	17000	1900	10176	1	1	1	1	0
99	2	17000	1100	10869	1	1	1	0	23000
100	5	27500	2300	16610	1	1	1	1	0
101	3	13500	1500	11304	1	1	1	1	0
102	2	27500	3000	23727	1	1	1	1	0
103	4	24970	2280	18602	1	1	1	1	0
104	2	3500	970	10065	1	1	0	0	0
105	2	17000	1169	10810	1	1	0	0	30000
106	2	13500	1800	20614	1	1	1	0	0
107	2	13500	728	4841	1	1	1	1	0
108	2	11000	1500	11235	1	1	1	1	0
109	3	17000	1500	9774	1	1	0	1	0
110	5	5500	900	12085	1	1	0	0	23000
111	3	17000	1500	17859	1	1	1	1	0
112	1	70711	2600	16661	1	1	1	1	0
113	3	7000	780	5692	1	1	1	0	20000
114	4	22500	1600	8191	1	1	1	1	0
115	2	13500	600	5086	0	1	1	0	2000
116	3	4500	1200	14178	1	1	1	0	1000
117	5	17000	900	8966	1	1	1	0	18000
118	4	13500	1500	11142	1	1	1	1	0
119	5	17000	2000	19555	1	1	1	1	0
120	3	23067	1740	10183	1	1	1	0	42000
121	1	17000	696	5974	1	0	0	0	0
122	1	2500	900	10111	1	1	0	0	0
123	2	7265	970	20437	1	1	0	0	0
124	2	10415	1500	9619	1	0	0	0	0
125	3	5500	750	16955	0	0	1	0	18000
126	2	4500	824	11647	1	1	0	0	0
127	1	22500	1900	11401	1	0	1	1	0
128	4	40000	2500	15205	1	1	1	1	0
129	2	4500	840	5984	1	1	1	1	0
130	1	22500	1800	18012	1	1	1	1	0
131	2	5500	1200	8447	1	1	0	0	1000
132	1	3689	576	12207	0	0	0	0	0
133	3	16356	1600	16227	0	1	1	0	28500
134	4	11000	1360	17045	1	1	0	0	0
135	3	5500	600	4644	0	1	0	0	9000

(Continued next page).

Table 1c. (Continued).

			Residence					Air Condition.	
t	Family Size	Income ($ per yr)	Size (SqFt)	Heat Loss (Btuh)	Elec. Range (1=yes)	Washer (1=yes)	Dryer (1=yes)	Central (1=yes)	Window (Btuh)
136	3	17000	2000	16731	1	1	1	1	2300
137	2	32070	6000	61737	1	1	1	1	0
138	2	27500	1250	7397	1	1	1	1	0
139	4	17000	840	5426	1	1	1	1	0
140	4	27500	3300	11023	1	1	1	1	0
141	2	11000	1200	10888	1	0	0	0	18000
142	1	.	1000	5446	1	0	0	0	0
143	3	36919	1200	8860	1	1	1	1	0
144	5	9000	720	5882	1	1	1	0	10000
145	5	21400	1300	6273	1	1	1	0	0
146	1	1500	375	6727	0	0	0	0	0
147	2	5063	1008	7195	1	0	0	0	0
148	1	3500	1650	13164	1	0	0	1	0
149	1	9488	850	9830	0	0	1	0	10000
150	1	27500	1200	8469	1	1	1	1	0
151	5	17000	1000	8006	0	1	1	0	16000
152	3	11000	2000	12608	1	1	1	1	0
153	7	22500	1225	11505	1	0	0	1	0
154	6	3500	1200	16682	1	1	0	0	0
155	3	9273	600	5078	1	1	0	0	15000
156	8	17000	1100	17912	1	0	0	0	0
157	3	17459	980	7984	0	1	1	1	0
158	5	11000	1200	14113	1	1	1	0	18000
159	3	9000	1600	21529	1	1	1	0	6000
160	2	11000	899	5731	0	1	1	0	28000
161	3	12068	1350	16331	1	1	1	0	6000
162	2	7000	672	8875	1	1	0	0	0
163	3	22500	1200	10424	1	1	0	0	23000
164	2	5500	1300	8636	1	1	1	1	0
165	2	12519	1000	24210	1	1	1	0	37000
166	2	29391	1400	12837	1	1	1	1	0
167	2	9000	400	4519	1	0	0	0	0
168	3	4664	1235	14274	1	1	0	0	6000
169	4	11000	720	6393	0	1	1	0	23000
170	.	.	.	.	.	.	.	.	.
171	3	18125	2300	16926	1	1	0	1	0
172	.	.	.	.	.	.	.	.	.
173	5	9000	720	6439	1	1	1	0	0
174	6	5500	1000	13651	1	1	0	0	0
175	5	14085	1400	14563	1	1	0	0	15000
176	2	9000	720	6540	0	1	1	1	0
177	6	17000	1470	8439	1	1	1	1	0
178	4	27500	1900	12345	1	1	1	1	18500
179	3	7000	480	3796	0	0	0	0	10000
180	3	13500	1300	7352	1	1	0	0	23000

(Continued next page).

Table 1c. (Continued).

t	Family Size	Income ($ per yr)	Residence Size (SqFt)	Residence Heat Loss (Btuh)	Elec. Range (1=yes)	Washer (1=yes)	Dryer (1=yes)	Air Condition. Central (1=yes)	Air Condition. Window (Btuh)
181	3	13437	1200	9502	1	1	1	1	0
182	3	14150	1300	8334	1	1	0	0	0
183	1	7000	1200	11941	1	1	0	0	21000
184	4	27500	1350	7585	1	1	1	1	0
185	2	32444	2900	15158	1	1	0	1	0
186	1	4274	400	7859	1	0	0	0	0
187	1	3500	600	14441	0	0	0	0	0
188	4	27500	2000	15462	1	1	1	1	0
189	4	40000	2900	13478	1	1	0	1	0
190	6	17000	5000	24132	1	0	1	1	0
191	1	2500	1400	17016	1	1	0	0	2000
192	7	9000	1400	13293	1	1	0	0	0
193	.	.	.	.	0	0	0	.	0
194	4	13500	780	5629	1	1	0	1	0
195	5	13500	1000	7281	1	1	1	1	0
196	2	13500	1169	11273	1	1	0	0	12000
197	2	40000	2400	13515	1	1	0	1	0
198	4	27500	1320	9865	1	1	1	0	29000
199	4	27500	1250	5759	1	1	1	1	0
200	1	3449	1200	18358	0	0	0	0	0
201	2	3500	425	4554	1	0	0	0	.
202	2	27500	1400	13496	1	0	0	1	0
203	4	7000	1300	11555	1	1	1	0	14000
204	2	3500	1800	23271	1	1	0	0	0
205	4	11000	720	5879	1	1	1	0	16000
206	7	9000	680	11528	1	0	0	0	0
207	4	14077	780	4829	1	1	1	0	10000
208	3	13500	2200	22223	1	1	1	0	24000
209	4	17000	1342	12050	1	1	1	1	0
210	4	3500	628	5369	1	1	1	0	24000
211	2	11000	920	5590	1	1	1	1	0
212	5	9000	1300	11510	1	1	1	0	19000
213	3	5500	1400	18584	1	1	1	0	23000
214	5	27500	2300	15480	1	1	1	1	0
215	3	20144	1700	11212	1	1	1	1	0
216	5	3500	1080	13857	0	0	0	0	0
217	2	22500	1800	17588	1	1	0	0	23000
218	6	22500	1900	15115	1	1	1	0	22000
219	5	6758	1200	16868	1	0	0	0	0
220	6	11000	2200	21884	1	1	1	1	0
221	3	17000	1500	11504	1	1	1	1	0
222	2	9000	600	5825	1	0	1	1	0
223	2	15100	1932	15760	1	1	1	0	0
224	1	7000	979	11700	1	1	1	0	1000

(Continued next page).

Table 1c. (Continued).

t	Type of Residence: Detached (1=yes)	Duplex or Apartment (1=yes)	Mobile Home (1=yes)	Elec. Water Heater (1=yes)	Freezer (kw)	Refrigerator (kw)
1	0	0	1	1	0	0.700
2	1	0	0	1	1.320	0.700
3	1	0	0	1	1.320	0.700
4	1	0	0	1	1.320	2.495
5	1	0	0	0	1.320	3.590
6	1	0	0	1	0	1.795
7	1	0	0	1	0	1.795
8	1	0	0	0	1.320	0.700
9	1	0	0	1	1.320	0.700
10	1	0	0	0	1.985	1.795
11	1	0	0	1	2.640	0.700
12	0	0	1	1	1.985	1.795
13	1	0	0	1	1.320	1.795
14	1	0	0	1	1.320	1.795
15	1	0	0	1	1.320	1.795
16	0	0	1	1	0	1.795
17	1	0	0	1	1.985	1.795
18	0	0	1	1	0	0.700
19	1	0	0	1	1.320	1.795
20	1	0	0	1	0	1.795
21	1	0	0	1	1.320	0.700
22	1	0	0	1	1.320	0.700
23	1	0	0	1	3.305	1.795
24	0	1	0	1	0	0.700
25	0	0	1	1	1.320	0.700
26	1	0	0	1	1.985	0.700
27	1	0	0	1	1.320	1.795
28	1	0	0	1	1.320	1.795
29	1	0	0	1	0	1.795
30	0	0	1	1	0	0.700
31	1	0	0	1	1.320	0.700
32	1	0	0	1	0	1.795
33	1	0	0	1	1.320	3.590
34	1	0	0	1	0	1.795
35	1	0	0	1	1.320	1.795
36	1	0	0	1	1.320	0.700
37	1	0	0	1	1.320	1.795
38	1	0	0	1	0	0.700
39	1	0	0	1	0	1.795
40	1	0	0	1	1.320	1.400
41	1	0	0	1	1.985	1.795
42	1	0	0	1	2.640	0.700
43	1	0	0	1	1.320	1.795
44	1	0	0	1	3.970	1.795
45	1	0	0	1	1.985	1.795

(Continued next page).

Table 1c. (Continued).

t	Type of Residence: Detached (1=yes)	Duplex or Apartment (1=yes)	Mobile Home (1=yes)	Elec. Water Heater (1=yes)	Freezer (kw)	Refrigerator (kw)
46	1	0	0	0	0	1.795
47	1	0	0	1	0	1.795
48	1	0	0	1	0	0.700
49	1	0	0	1	0	2.495
50	1	0	0	1	1.985	1.795
51	0	0	1	1	1.985	0.700
52	1	0	0	0	1.320	1.795
53	0	0	1	1	0	0.700
54	1	0	0	1	1.320	1.795
55	1	0	0	1	1.320	1.795
56	1	0	0	1	1.320	1.795
57	1	0	0	1	1.320	1.795
58	1	0	0	1	1.320	1.795
59	1	0	0	1	1.320	1.795
60	1	0	0	0	0	0.700
61	1	0	0	1	1.320	0.700
62	1	0	0	1	1.320	0.700
63	1	0	0	1	1.320	0.700
64	0	1	0	1	0	1.795
65	0	0	1	1	0	0.700
66	0	0	1	1	0	1.795
67	1	0	0	1	1.320	1.795
68	1	0	0	1	1.320	0.700
69	1	0	0	1	0	1.400
70	1	0	0	1	1.320	1.795
71	1	0	0	1	1.985	1.795
72	1	0	0	0	1.320	1.795
73	1	0	0	0	0	1.795
74	1	0	0	0	0	0.700
75	1	0	0	0	1.320	0.700
76	1	0	0	1	0	0.700
77	0	1	0	0	1.320	0.700
78	1	0	0	1	0	1.795
79	1	0	0	0	1.320	1.795
80	1	0	0	1	1.320	1.795
81	0	0	1	1	1.320	2.495
82	1	0	0	1	1.985	1.795
83	1	0	0	1	1.320	0.700
84	1	0	0	1	1.985	1.795
85	1	0	0	1	0	1.795
86	1	0	0	1	1.320	2.495
87	1	0	0	0	0	1.795
88	1	0	0	0	3.305	0.700
89	1	0	0	1	1.985	0.700
90	1	0	0	1	1.985	1.795

(Continued next page).

Table 1c. (Continued).

t	Type of Residence: Detached (1=yes)	Duplex or Apartment (1=yes)	Mobile Home (1=yes)	Elec. Water Heater (1=yes)	Freezer (kw)	Refrigerator (kw)
91	1	0	0	1	0	1.795
92	1	0	0	1	1.985	1.795
93	1	0	0	1	1.320	1.795
94	1	0	0	1	0	1.795
95	0	0	1	1	0	1.795
96	1	0	0	1	1.985	0.700
97	1	0	0	0	1.320	0.700
98	1	0	0	1	1.320	0.700
99	1	0	0	1	2.640	1.795
100	1	0	0	1	1.320	1.795
101	1	0	0	1	1.320	1.795
102	1	0	0	1	0	2.495
103	1	0	0	1	1.320	1.795
104	1	0	0	1	1.320	0.700
105	1	0	0	1	1.320	0.700
106	1	0	0	1	0	1.795
107	1	0	0	1	1.320	0.700
108	1	0	0	1	1.320	0.700
109	1	0	0	1	0	1.795
110	1	0	0	1	1.320	0.700
111	1	0	0	1	1.320	1.795
112	1	0	0	1	3.970	1.795
113	0	0	1	1	0	1.795
114	1	0	0	1	1.320	1.795
115	0	0	1	1	0	0.700
116	1	0	0	1	0	1.795
117	1	0	0	1	1.320	1.795
118	1	0	0	1	1.985	0.700
119	1	0	0	1	0	0.700
120	0	0	1	1	1.320	1.795
121	0	0	1	1	0	1.795
122	1	0	0	1	0	0.700
123	1	0	0	1	0	0.700
124	1	0	0	1	0	1.795
125	1	0	0	0	1.320	1.795
126	1	0	0	1	0	0.700
127	1	0	0	0	1.320	1.795
128	1	0	0	1	0	2.495
129	0	0	1	1	1.320	1.795
130	1	0	0	1	0	1.795
131	1	0	0	1	0	0.700
132	1	0	0	0	1.320	1.795
133	1	0	0	1	1.320	0.700
134	1	0	0	1	1.320	0.700
135	0	0	1	1	0	0.700

(Continued next page).

Table 1c. (Continued).

t	Type of Residence: Detached (1=yes)	Duplex or Apartment (1=yes)	Mobile Home (1=yes)	Elec. Water Heater (1=yes)	Freezer (kw)	Refrigerator (kw)
136	1	0	0	1	1.985	1.795
137	1	0	0	1	1.985	1.400
138	0	1	0	1	0	1.795
139	0	0	1	1	0	0.700
140	0	1	0	1	7.265	1.795
141	1	0	0	1	1.320	0.700
142	0	1	0	1	0	1.795
143	0	1	0	1	0	0.700
144	0	0	1	1	0	0.700
145	0	0	1	1	0	1.795
146	1	0	0	0	0	1.795
147	1	0	0	0	0	0.700
148	1	0	0	0	0	1.795
149	1	0	0	0	1.320	1.795
150	0	1	0	1	0	1.795
151	1	0	0	1	2.640	1.795
152	1	0	0	1	0	1.795
153	1	0	0	0	1.320	1.795
154	1	0	0	1	1.320	0.700
155	0	0	1	1	0	0.700
156	1	0	0	1	1.320	0.700
157	0	0	1	1	1.320	1.795
158	1	0	0	1	3.970	2.495
159	1	0	0	1	0	0.700
160	1	0	0	1	1.320	1.795
161	1	0	0	1	1.320	1.795
162	1	0	0	1	1.320	0.700
163	1	0	0	1	0	1.795
164	1	0	0	1	0	1.795
165	1	0	0	1	2.640	3.590
166	1	0	0	1	1.320	0.700
167	0	1	0	1	0	0.700
168	1	0	0	1	2.640	0.700
169	0	0	1	1	0	1.795
170	.	.	.	.	.	.
171	1	0	0	1	1.320	1.795
172	.	.	.	.	.	.
173	0	0	1	1	1.320	0.700
174	1	0	0	1	1.320	0.700
175	1	0	0	1	1.985	1.795
176	0	0	1	1	0	0.700
177	1	0	0	1	3.970	1.795
178	1	0	0	1	1.985	1.795
179	0	0	1	0	0	0.700
180	1	0	0	1	1.320	1.795

(Continued next page).

Table 1c. (Continued).

t	Type of Residence: Detached (1=yes)	Duplex or Apartment (1=yes)	Mobile Home (1=yes)	Elec. Water Heater (1=yes)	Freezer (kw)	Refrigerator (kw)
181	1	0	0	1	1.985	0.700
182	1	0	0	1	1.985	1.795
183	1	0	0	1	1.320	1.795
184	1	0	0	1	1.320	1.795
185	1	0	0	1	0	2.495
186	1	0	0	1	0	1.795
187	1	0	0	0	1.320	0
188	1	0	0	1	1.320	1.795
189	1	0	0	1	0	1.795
190	1	0	0	0	1.985	2.495
191	1	0	0	1	0	1.795
192	1	0	0	1	0	1.795
193	0	0	1	0	0	0
194	0	0	1	1	1.320	1.795
195	1	0	0	1	1.320	0.700
196	1	0	0	1	1.985	1.795
197	1	0	0	1	1.985	1.795
198	0	0	1	1	0	1.795
199	1	0	0	1	1.320	1.795
200	1	0	0	0	0	0.700
201	0	1	0	1	1.985	0
202	0	1	0	1	0	1.795
203	1	0	0	1	1.320	0.700
204	1	0	0	1	0	0.700
205	0	0	1	1	1.320	0.700
206	1	0	0	0	1.320	0.700
207	0	0	1	1	0	0.700
208	1	0	0	1	1.320	2.495
209	1	0	0	1	1.320	1.795
210	1	0	0	1	0	0.700
211	0	1	0	1	0	0.700
212	1	0	0	1	0	1.795
213	1	0	0	1	1.320	1.795
214	1	0	0	1	1.985	1.795
215	1	0	0	1	1.320	1.795
216	1	0	0	0	0	0.700
217	1	0	0	1	1.320	1.795
218	1	0	0	1	1.985	1.795
219	1	0	0	0	0	0.700
220	1	0	0	1	1.320	1.795
221	1	0	0	1	1.320	1.795
222	0	1	0	1	0	1.795
223	1	0	0	1	1.320	1.795
224	1	0	0	1	0	1.795

Source: Gallant and Koenker(1984).

2. LEAST SQUARES ESTIMATORS AND MATTERS OF NOTATION

Univariate responses $y_{\alpha t}$ for $t = 1, 2, \ldots, n$ and $\alpha = 1, 2, \ldots, M$ are presumed to be related to k-dimensional input vectors x_t as follows

$$y_{\alpha t} = f_\alpha(x_t, \theta_\alpha^\circ) + e_{t\alpha} \qquad \alpha = 1, 2, \ldots, M\,; \; t = 1, 2, \ldots, n$$

where each $f_\alpha(x, \theta_\alpha)$ is a known function, each θ_α° is a p_α-dimensional vector of unknown parameters, and the $e_{\alpha t}$ represent unobservable observational or experimental errors. As previously, we write θ_α° to emphasize that it is the true, but unknown, value of the parameter vector θ_α that is meant; θ_α itself is used to denote instances when the parameter vector is treated as a variable. Writing

$$e_t = \begin{pmatrix} e_{1t} \\ e_{2t} \\ \vdots \\ e_{Mt} \end{pmatrix} \quad ,$$

the error vectors e_t are assumed to be independently and identically distributed with mean zero and unknown variance-covariance matrix Σ ,

$$\Sigma = C(e_t, e_t') \qquad t = 1, 2, \ldots, n\,,$$

whence

$$C(e_{\alpha t}, e_{\beta t}) = \begin{cases} \sigma_{\alpha\beta} & t = s \\ 0 & t \neq s \end{cases}$$

with $\sigma_{\alpha\beta}$ denoting the elements of Σ .

In the literature one finds two conventions for writing this model in a vector form. One emphasizes the fact that the model consists of M separate univariate nonlinear regressions

$$y_\alpha = f_\alpha(\theta_\alpha^\circ) + e_\alpha \qquad \alpha = 1, 2, \ldots, M$$

with y_α being an n-vector as described below and the other emphasizes the multivariate nature of the data

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

with y_t being an M-vector; Simply to have labels to distinguish the two, we shall follow Zellner (1962) and refer to the first notational scheme as the "seemingly unrelated" (nonlinear regressions) structure the second as the multivariate (nonlinear regression) structure. Let us take these up in turn.

The "seemingly unrelated" notational scheme follows the same conventions used in Chapter 1. Write

$$y_\alpha = {}_n\begin{pmatrix} y_{\alpha 1} \\ y_{\alpha 2} \\ \vdots \\ y_{\alpha n} \end{pmatrix}_1$$

$$f_\alpha(\theta_\alpha) = {}_n\begin{pmatrix} f_\alpha(x_1, \theta_\alpha) \\ f_\alpha(x_2, \theta_\alpha) \\ \vdots \\ f_\alpha(x_n, \theta_\alpha) \end{pmatrix}_1$$

$$e_\alpha = {}_n\begin{pmatrix} e_{\alpha 1} \\ e_{\alpha 2} \\ \vdots \\ e_{\alpha n} \end{pmatrix}_1 \quad .$$

In this notation, each regression is written as

$$y_\alpha = f_\alpha(\theta_\alpha^\circ) + e_\alpha \qquad \alpha = 1, 2, \ldots, M$$

with (Problem 1)

$$C(e_\alpha, e'_\beta) = \sigma_{\alpha\beta}\ {}_nI_n \ .$$

Denote the Jacobian of $f_\alpha(\theta_\alpha)$ by

$$F_\alpha(\theta_\alpha) = (\partial/\partial\theta'_\alpha)\ f_\alpha(\theta_\alpha)$$

which is of order n by p_α. Illustrating with Example 1 we have:

EXAMPLE 1 (continued). The independent variables are the logarithms of expenditure normalized prices. From Tables 1a and 1b we obtain a few instances

$$x_1 = \ell n[(3.90,\ 2.86,\ 1.06)/(0.46931)]' = (2.11747,\ 1.80731,\ 0.81476)'$$
$$x_2 = \ell n[(3.90,\ 2.86,\ 1.06)/(0.79539)]' = (1.58990,\ 1.27974,\ 0.28719)'$$
$$\vdots$$
$$x_{20} = \ell n[(3.90,\ 2.86,\ 1.06)/(1.37160)]' = (1.04500,\ 0.73484,\ -0.25771)'$$
$$x_{21} = \ell n[(3.90,\ 2.86,\ 1.78)/(0.92766)]' = (1.43607,\ 1.12591,\ 0.65170)'$$
$$\vdots$$
$$x_{40} = \ell n[(3.90,\ 2.86,\ 1.78)/(2.52983)]' = (0.43282,\ 0.12267,\ -0.35154)'$$
$$x_{41} = \ell n[(3.90,\ 3.90,\ 1.06)/(1.14741)]' = (1.22347,\ 1.22347,\ -0.079238)'$$
$$\vdots$$
$$x_{224} = \ell n[(6.56,\ 3.90,\ 1.78)/(1.15897)]' = (1.73346,\ 1.21344,\ 0.42908)' \ .$$

The vectors of dependent variables are for $\alpha = 1$

$$\underset{224}{y_\alpha} = \underset{224}{\begin{pmatrix} \ell n(0.662888/0.056731) \\ \ell n(0.644427/0.103444) \\ \vdots \\ \ell n(0.521465/0.179133) \end{pmatrix}}_{1} = \underset{224}{\begin{pmatrix} 2.45829 \\ 1.82933 \\ \vdots \\ 1.06851 \end{pmatrix}}_{1} \ ,$$

and for $\alpha = 2$

$$y_\alpha = \underset{224}{\begin{pmatrix} \ell n(0.280382/0.056731) \\ \ell n(0.252128/0.103444) \\ \vdots \\ \ell n(0.299403/0.179133) \end{pmatrix}}_{1} = \underset{224}{\begin{pmatrix} 1.59783 \\ 0.89091 \\ \vdots \\ 0.51366 \end{pmatrix}}_{1} .$$

Recall that

$$y_{1t} = \ell n[(a_1 + x_t'b_{(1)})/(a_3 + x_t'b_{(3)})] + e_{1t}$$

$$y_{2t} = \ell n[(a_2 + x_t'b_{(2)})/(a_3 + x_3'b_{(3)})] + e_{2t} ,$$

with $b_{(\alpha)}$ denoting the α-th row of

$$B = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}$$

and with $a' = (a_1, a_2, a_3)$. Note that if both a and B are multiplied by some common factor δ to obtain $\bar{a} = \delta a$ and $\bar{B} = \delta B$ we shall have

$$(a_\alpha + x'b_{(\alpha)})/(a_3 + x'b_{(3)}) = (\bar{a}_\alpha + x'\bar{b}_{(\alpha)})/(\bar{a}_3 + x'\bar{b}_{(3)}) .$$

Thus the parameters of the model can only be determined to within a scalar multiple. In order to estimate the model it is necessary to impose a normalization rule. Our choice is to set $a_3 = -1$. With this choice we write the model as

$$y_{1t} = f_1(x_t, \theta_1^\circ) + e_{1t}$$

$$y_{2t} = f_2(x_t, \theta_2^\circ) + e_{2t}$$

with

$$f_\alpha(x,\theta_\alpha) = \ln[(a_\alpha + b'_{(\alpha)}x)/(-1 + b'_{(3)}x)] \qquad \alpha = 1, 2$$

$$\theta_1' = (a_1, b_{11}, b_{12}, b_{13}, b_{31}, b_{32}, b_{33})$$

$$\theta_2' = (a_2, b_{21}, b_{22}, b_{23}, b_{31}, b_{32}, b_{33}) \ . \ \square$$

Recognizing that what we have is M instances of the univariate nonlinear regression model of Chapter 1, we can apply our previous results and estimate the parameters θ_α° of each model by computing $\hat{\theta}_\alpha^\#$ to minimize

$$SSE_\alpha(\theta_\alpha) = [y_\alpha - f_\alpha(\theta_\alpha)]'[y_\alpha - f_\alpha(\theta_\alpha)]$$

for $\alpha = 1, 2, \ldots, M$. This done, the elements $\sigma_{\alpha\beta}$ of Σ can be estimated by

$$\hat{\sigma}_{\alpha\beta} = [y_\alpha - f_\alpha(\hat{\theta}_\alpha^\#)]'[y_\beta - f_\beta(\hat{\theta}_\beta^\#)]/n \qquad \alpha,\beta = 1, 2, \ldots, M \ .$$

Let $\hat{\Sigma}$ denote the M by M matrix with typical element $\hat{\sigma}_{\alpha\beta}$. Equivalently, if we write

$$\hat{e}_\alpha = y_\alpha - f_\alpha(\hat{\theta}_\alpha^\#) \qquad \alpha = 1, 2, \ldots, M$$

$$\hat{E} = [\hat{e}_1 \vdots \hat{e}_2 \vdots \ldots \vdots \hat{e}_M]$$

then

$$\hat{\Sigma} = (1/n)\hat{E}'\hat{E} \ .$$

We illustrate with Example 1.

EXAMPLE 1 (continued). Fitting

$$y_1 = f_1(\theta_1) + e_1$$

by the methods of Chapter 1 we have from Figure 1a that

Figure 1a. First Equation of Example 1 Fitted by the Modified Gauss-Newton Method.

SAS Statements:

```
PROC NLIN DATA=EXAMPLE1 METHOD=GAUSS ITER=50 CONVERGENCE=1.E-13;
PARMS B11=0 B12=0 B13=0 B31=0 B32=0 B33=0 A1=-9; A3=-1;
PEAK=A1+B11*X1+B12*X2+B13*X3;    BASE=A3+B31*X1+B32*X2+B33*X3;
MODEL Y1=LOG(PEAK/BASE);
DER.A1 =1/PEAK;
DER.B11=1/PEAK*X1;  DER.B31=-1/BASE*X1;
DER.B12=1/PEAK*X2;  DER.B32=-1/BASE*X2;
DER.B13=1/PEAK*X3;  DER.B33=-1/BASE*X3;
OUTPUT OUT=WORK02 RESIDUAL=E1;
```

Output:

```
           S T A T I S T I C A L   A N A L Y S I S   S Y S T E M            1

                  NON-LINEAR LEAST SQUARES ITERATIVE PHASE

            DEPENDENT VARIABLE: Y1          METHOD: GAUSS-NEWTON

  ITERATION              B11              B12              B13        RESIDUAL SS
                         B31              B32              B33
                          A1

      0       0.000000E+00     0.000000E+00     0.000000E+00      72.21326991
              0.000000E+00     0.000000E+00     0.000000E+00
               -9.00000000

      .
      .
      .

     16        -0.83862780      -1.44241315       2.01535561      36.50071896
                0.46865734      -0.19468166      -0.38299626
               -1.98254583

NOTE: CONVERGENCE CRITERION MET.

           S T A T I S T I C A L   A N A L Y S I S   S Y S T E M            3

    NON-LINEAR LEAST SQUARES SUMMARY STATISTICS     DEPENDENT VARIABLE Y1

      SOURCE                  DF      SUM OF SQUARES       MEAN SQUARE

      REGRESSION               7      1019.72335676      145.67476525
      RESIDUAL               217        36.50071896        0.16820608
      UNCORRECTED TOTAL      224      1056.22407572

      (CORRECTED TOTAL)      223        70.01946051

PARAMETER        ESTIMATE        ASYMPTOTIC              ASYMPTOTIC 95 %
                                 STD. ERROR           CONFIDENCE INTERVAL
                                                     LOWER             UPPER
B11           -0.83862780        1.37155782     -3.54194099        1.86468538
B12           -1.44241315        1.87671707     -5.14138517        2.25655887
B13            2.01535561        1.44501283     -0.83273595        4.86344716
B31            0.46865734        0.12655505      0.21921985        0.71809482
B32           -0.19468166        0.21864114     -0.62561901        0.23625569
B33           -0.38299626        0.09376286     -0.56780098       -0.19819153
A1            -1.98254583        1.03138455     -4.01538427        0.05029260
```

$$\hat{\theta}_1^{\#} = \begin{pmatrix} \hat{a}_1 \\ \hat{b}_{11} \\ \hat{b}_{12} \\ \hat{b}_{13} \\ \hat{b}_{31} \\ \hat{b}_{32} \\ \hat{b}_{33} \end{pmatrix} = \begin{pmatrix} -1.98254583 \\ -0.83862780 \\ -1.44241315 \\ 2.01535561 \\ 0.46865734 \\ -0.19468166 \\ -0.38299626 \end{pmatrix}$$

and from Figure 1b that

$$\hat{\theta}_2^{\#} = \begin{pmatrix} \hat{a}_2 \\ \hat{b}_{21} \\ \hat{b}_{22} \\ \hat{b}_{23} \\ \hat{b}_{31} \\ \hat{b}_{32} \\ \hat{b}_{33} \end{pmatrix} = \begin{pmatrix} -1.11401781 \\ 0.41684196 \\ -1.30951752 \\ 0.73956410 \\ 0.24777391 \\ 0.07675306 \\ -0.39514717 \end{pmatrix}$$

Some aspects of these computations deserve comment. In this instance, the convergence of the modified Gauss-Newton method is fairly robust to the choice of starting values so we have taken the simple expedient of starting with a value ${}_0\theta_\alpha$ with $f_\alpha(x, {}_0\theta_\alpha) \doteq \bar{y}_\alpha$. The first full step away from ${}_0\theta_\alpha$

$${}_1\theta_\alpha = {}_0\theta_\alpha + [F_\alpha'({}_0\theta_\alpha)\ F_\alpha({}_0\theta_\alpha)]^{-1}\ F_\alpha'({}_0\theta_\alpha)[y_\alpha - f_\alpha({}_0\theta_\alpha)]$$

is such that

$$({}_1a_\alpha + {}_1b_{(\alpha)}'x)/(-1 + {}_1b_{(3)}'x)$$

is negative for some of the x_t ; this results in an error condition when taking logarithms. Obviously one need only take care to choose a step length ${}_0\lambda_\alpha$ small enough such that

Figure 1b. Second Equation of Example 1 Fitted by the Modified Gauss-Newton Method.

SAS Statements:

```
PROC NLIN DATA=EXAMPLE1 METHOD=GAUSS ITER=50 CONVERGENCE=1.E-13;
PARMS B21=0 B22=0 B23=0 B31=0 B32=0 B33=0 A2=-3; A3=-1;
INTER=A2+B21*X1+B22*X2+B23*X3;  BASE=A3+B31*X1+B32*X2+B33*X3;
MODEL Y2=LOG(INTER/BASE);
DER.A2 =1/INTER;
DER.B21=1/INTER*X1;  DER.B31=-1/BASE*X1;
DER.B22=1/INTER*X2;  DER.B32=-1/BASE*X2;
DER.B23=1/INTER*X3;  DER.B33=-1/BASE*X3;
OUTPUT OUT=WORK03 RESIDUAL=E2;
```

Output:

```
                 S T A T I S T I C A L   A N A L Y S I S   S Y S T E M              4

                       NON-LINEAR LEAST SQUARES ITERATIVE PHASE

                   DEPENDENT VARIABLE: Y2          METHOD: GAUSS-NEWTON

   ITERATION             B21             B22             B23         RESIDUAL SS
                         B31             B32             B33
                          A2

       0        0.000000E+00    0.000000E+00    0.000000E+00         37.16988980
                0.000000E+00    0.000000E+00    0.000000E+00
                 -3.00000000
       .
       .
       .

      16          0.41684196     -1.30951752      0.73956410         19.70439405
                  0.24777391      0.07675306     -0.39514717
                 -1.11401781

NOTE: CONVERGENCE CRITERION MET.

                 S T A T I S T I C A L   A N A L Y S I S   S Y S T E M              6

     NON-LINEAR LEAST SQUARES SUMMARY STATISTICS     DEPENDENT VARIABLE Y2

       SOURCE                   DF     SUM OF SQUARES         MEAN SQUARE

       REGRESSION                7       265.36865902         37.90980843
       RESIDUAL                217        19.70439405          0.09080366
       UNCORRECTED TOTAL       224       285.07305307

       (CORRECTED TOTAL)       223        36.70369496

PARAMETER         ESTIMATE         ASYMPTOTIC                ASYMPTOTIC 95 %
                                   STD. ERROR              CONFIDENCE INTERVAL
                                                         LOWER           UPPER
B21             0.41684196         0.44396622      -0.45820663      1.29189056
B22            -1.30951752         0.60897020      -2.50978567     -0.10924936
B23             0.73956410         0.54937638      -0.34324582      1.82237401
B31             0.24777391         0.13857700      -0.02535860      0.52090642
B32             0.07675306         0.18207332      -0.28210983      0.43561595
B33            -0.39514717         0.08932410      -0.57120320     -0.21909114
A2             -1.11401781         0.34304923      -1.79016103     -0.43787460
```

Figure 1c. Contemporaneous Variance-Covariance Matrix of Example 1 Estimated from Single Equation Residuals.

SAS Statements:

```
DATA WORK04;  MERGE WORK02 WORK03;  KEEP T E1 E2;
PROC MATRIX FW=20; FETCH E DATA=WORK04(KEEP=E1 E2);
SIGMA=E'*E#/224;  PRINT SIGMA;  P=HALF(INV(SIGMA));  PRINT P;
```

Output:

SIGMA	COL1	COL2
ROW1	0.1629496382006	0.09015433203941
ROW2	0.09015433203941	0.08796604486025

P	COL1	COL2
ROW1	3.764814163903	-3.85846955764
ROW2	0	3.371649857133

$$ {}_1\theta_\alpha = {}_0\theta_\alpha + {}_0\lambda_\alpha [F'_\alpha({}_0\theta_\alpha)\, F_\alpha({}_0\theta_\alpha)]^{-1} F'({}_0\theta_\alpha)[y_\alpha - f_\alpha({}_0\theta_\alpha)] $$

is in range to avoid this difficulty. Thus, this situation is not a problem for properly written code. Other than cluttering up the output (suppressed in the figures), the SAS code seems to behave reasonably well. **See Problem 7 for another approach to this problem.**

Lastly, we compute

$$ \hat{\Sigma} = \begin{pmatrix} 0.1629496382006 & 0.09015433203941 \\ 0.09015433203941 & 0.08796604486025 \end{pmatrix} $$

as shown in Figure 1c. For later use we compute

$$ \hat{P} = \begin{pmatrix} 3.764814163903 & -3.85846955764 \\ 0 & 3.371659857133 \end{pmatrix} $$

with $\hat{\Sigma}^{-1} = \hat{P}'\hat{P}$. $\square$

The set of M regressions can be arranged in a single regression

$$ y = f(\theta^\circ) + e $$

by writing

$$ \underset{nM\,1}{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{pmatrix} $$

$$ \underset{nM\,1}{f(\theta)} = \begin{pmatrix} f_1(\theta_1) \\ f_2(\theta_2) \\ \vdots \\ f_M(\theta_M) \end{pmatrix} $$

$$ \underset{nM\,1}{e} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_M \end{pmatrix} $$

$$\underset{p}{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_M \end{pmatrix}_{1}$$

with $p = \Sigma_{\alpha=1}^{M} p_\alpha$. In order to work out the variance-covariance matrix of e let us review Kroneker product notation.

If A is an k by ℓ matrix and B is m by n then their Kroneker product, denoted as $A \otimes B$ is the km by ℓn matrix

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1\ell}B \\ a_{21}B & a_{22}B & \cdots & a_{2\ell}B \\ \vdots & \vdots & & \\ a_{k1}B & a_{k2}B & \cdots & a_{k\ell}B \end{pmatrix}$$

The operations of matrix transposition and Kroneker product formation commute; viz.

$$(A \otimes B)' = (A' \otimes B') .$$

If A and C are conformable for multiplication, that is, C has as many rows as A has columns, and B and D are conformable as well then

$$(A \otimes B)(C \otimes D) = (AC \otimes BD) .$$

It follows immediately that if both A and B are square and invertable then

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$$

that is, inversion and Kroneker product formation commute.

In this notation, the variance-covariance matrix of the errors is

$$C(e,e') = \begin{pmatrix} C(e_1, e_1') & C(e_1,e_2') & \cdots & C(e_1,e_M') \\ C(e_2, e_1') & C(e_2,e_2') & \cdots & C(e_2,e_M') \\ \vdots & & & \\ C(e_M, e_1') & C(e_M,e_2') & & C(e_M,e_M') \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_{11} I & \sigma_{12} I & \cdots & \sigma_{1M} I \\ \sigma_{21} I & \sigma_{22} I & \cdots & \sigma_{2M} I \\ \vdots & \vdots & & \vdots \\ \sigma_{M1} I & \sigma_{M2} I & \cdots & \sigma_{MM} I \end{pmatrix}$$

$$= \quad \Sigma \otimes I \,;$$

the identity is n by n while Σ is M by M so the resultant $\Sigma \otimes I$ is nM by nM .

Factor Σ^{-1} as $\Sigma^{-1} = P'P$ and consider the rotated model

$$(P \otimes I)'y = (P \otimes I)'f(\theta) + (P \otimes I)'e$$

or

$$"y" = "f"(\theta) + "e" \;.$$

Since

$$C("e", "e"') = (P \otimes I)'(\Sigma \otimes I)(P \otimes I)$$

$$= (P'\Sigma P) \otimes I$$

$$= [P'(P')^{-1}P^{-1}P] \otimes I$$

$$= {}_M I_M \otimes {}_n I_n$$

$$= {}_{nM} I_{nM}$$

the model

$$"y" = "f"(\theta^\circ) + "e"$$

is simply a univariate nonlinear model and θ° can be estimated by minimizing

$$\begin{aligned} S(\theta,\Sigma) &= ["y" - "f"(\theta)]'["y" - "f"(\theta)] \\ &= [y - f(\theta)]'(P \otimes I)'(P \otimes I)[y - f(\theta)] \\ &= [y - f(\theta)]'(\Sigma^{-1} \otimes I)[y - f(\theta)] \ . \end{aligned}$$

Of course Σ is unknown so one adopts the obvious expedient (Problem 4) of replacing Σ by $\hat{\Sigma}$ and estimating θ° by

$$\hat{\theta} \text{ minimizing } S(\theta,\hat{\Sigma}) \ .$$

These ideas are easier to implement if we adopt the multivariate notational scheme rather than the "seemingly unrelated" regressions scheme. Accordingly, let

$$\underset{M \times 1}{y_t} = \begin{pmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{Mt} \end{pmatrix} \qquad t = 1, 2, \ldots, n$$

$$\underset{M \times 1}{f(x,\theta)} = \begin{pmatrix} f_1(x,\theta_1) \\ f_2(x,\theta_2) \\ \vdots \\ f_M(x,\theta_M) \end{pmatrix}$$

$$\underset{M \times 1}{e_t} = \begin{pmatrix} e_{1t} \\ e_{2t} \\ \vdots \\ e_{Mt} \end{pmatrix} \qquad t = 1, 2, \ldots, n$$

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_M \end{pmatrix}_{p \times 1} \qquad\qquad p = \Sigma_{\alpha=1}^{M} p_\alpha$$

whence the model may be written as the multivariate nonlinear regression

$$y_t = f(x_t, \theta^\circ) + e_t \qquad\qquad t = 1, 2, \ldots, n .$$

In this scheme,

$$S(\theta,\Sigma) = \Sigma_{t=1}^{n} [y_t - f(x_t,\theta)]' \Sigma^{-1} [y_t - f(x_t,\theta)] .$$

To see that this is so, let $\sigma^{\alpha\beta}$ denote the elements of Σ^{-1} and write

$$\begin{aligned} S(\theta,\Sigma) &= \Sigma_{t=1}^{n} [y_t - f(x_t,\theta)]' \Sigma^{-1} [y_t - f(x_t,\theta)] \\ &= \Sigma_{t=1}^{n} \Sigma_{\alpha=1}^{M} \Sigma_{\beta=1}^{M} \sigma^{\alpha\beta} [y_{\alpha t} - f_\alpha(x_t,\theta_\alpha)] [y_{\beta t} - f_\beta(x_t,\theta_\beta)] \\ &= \Sigma_{\alpha=1}^{M} \Sigma_{\beta=1}^{M} \sigma^{\alpha\beta} [y_\alpha - f(\theta_\alpha)]' [y_\beta - f(\theta_\beta)] \\ &= [y - f(\theta)]' (\Sigma^{-1} \otimes I) [y - f(\theta)] . \end{aligned}$$

The advantage of the multivariate notational scheme in writing code derives from the fact that it is natural to group observations (y_t, x_t) on the same subject together and process them serially for $t = 1, 2, \ldots, n$. With $S(\theta,\Sigma)$ written as

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

one can see at sight that it suffices to fetch (y_t, x_t), compute $[y_t - f(x_t,\theta)]' \Sigma^{-1} [y_t - f(x_t,\theta)]$, add the result to an accumulator and continue.

The notation is also suggestive of a transformation that permits the use of univariate nonlinear regression programs for multivariate computations. Observe that if Σ^{-1} factors as $\Sigma^{-1} = P'P$ then

$$S(\theta,\Sigma) = \Sigma_{t=1}^{n}[Py_t - Pf(x_t,\theta)]'[Py_t - Pf(x_t,\theta)] \quad .$$

Writing $p'_{(\alpha)}$ to denote the α-th row of P we have

$$S(\theta,\Sigma) = \Sigma_{t=1}^{n} \Sigma_{\alpha=1}^{M}[p'_{(\alpha)}y_t - p'_{(\alpha)}f(x_t,\theta)]^2 \ .$$

One now has $S(\theta,\Sigma)$ expressed as the sum of squares of univariate entities, what remains is to find a notational scheme to remove the double summation. To this end, put

$$s = M(t-1) + \alpha$$

$$"y_s" = p'_{(\alpha)}y_t$$

$$"x_s" = (p'_{\alpha},x'_t)'$$

$$"f_s"("x_s",\theta) = p'_{(\alpha)}f(x_t,\theta)$$

for $\alpha = 1, 2, \ldots, M$ and $t = 1, 2, \ldots, n$ whence

$$S(\theta,\Sigma) = \Sigma_{s=1}^{nM}["y_s" - "f"("x_s",\theta)]^2 \ .$$

We illustrate these ideas with the example.

EXAMPLE 1 (continued). Recall that the model is

$$y_{1t} = f_1(x_t,\theta_1) + e_{t1}$$

$$y_{2t} = f_2(x_t,\theta_2) + e_{t2}$$

with

$$f_\alpha(x,\theta_\alpha) = \ln[(a_\alpha + b'_{(\alpha)})/(-1 + b'_{(3)}x)] \qquad \alpha = 1, 2$$

$$\theta'_1 = (a_1, b_{11}, b_{12}, b_{13}, b_{31}, b_{32}, b_{33})$$

$$\theta'_2 = (a_2, b_{21}, b_{22}, b_{23}, b_{31}, b_{32}, b_{33}) \ .$$

As the model is written, the notation suggests that $b_{(3)}$ is the same for both $\alpha = 1$ and $\alpha = 2$ which up to now has not been the case. To have a notation that reflects this fact write

$$f_\alpha(x,\theta_\alpha) = \ln[(a_\alpha + b'_{(\alpha)})/(-1 + b'_{\alpha(3)}x)] \qquad \alpha = 1, 2$$

$$\theta'_1 = (a_1, b_{11}, b_{12}, b_{13}, b_{131}, b_{132}, b_{133})$$

$$\theta'_2 = (a_2, b_{21}, b_{22}, b_{23}, b_{231}, b_{232}, b_{233})$$

to emphasize the fact that the equality constraint is not imposed. The multivariate model is, then,

$$y_t = f(x_t,\theta) + e_t$$

with

$$f(x_t,\theta) = \begin{pmatrix} \ln \dfrac{a_1 + b_{11}x_{1t} + b_{12}x_{2t} + b_{13}x_{3t}}{-1 + b_{131}x_{1t} + b_{132}x_{2t} + b_{133}x_{3t}} \\ \\ \ln \dfrac{a_2 + b_{21}x_{2t} + b_{22}x_{2t} + b_{23}x_{3t}}{-1 + b_{231}x_{1t} + b_{232}x_{2t} + b_{233}x_{3t}} \end{pmatrix}$$

$$\theta = \begin{pmatrix} a_1 \\ b_{11} \\ b_{12} \\ b_{13} \\ b_{131} \\ b_{132} \\ b_{133} \\ a_2 \\ b_{21} \\ b_{22} \\ b_{23} \\ b_{231} \\ b_{232} \\ b_{233} \end{pmatrix} ,$$

and $y_t = \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix}$, $e_t = \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}$, x_t as before. To illustrate, from Table 1a for t = 1 we have

$$y_t = \begin{pmatrix} \ell n(0.662888/0.056731) \\ \ell n(0.280382/0.056731) \end{pmatrix} = \begin{pmatrix} 2.45829 \\ 1.59783 \end{pmatrix}$$

and for t = 2

$$y_t = \begin{pmatrix} \ell n(0.644427/0.103444) \\ \ell n(0.252128/0.013444) \end{pmatrix} = \begin{pmatrix} 1.82933 \\ 0.89091 \end{pmatrix} ;$$

as previously from Tables 1a and 1b we have

$$x_1 = \begin{pmatrix} 2.11747 \\ 1.80731 \\ 0.81476 \end{pmatrix} , \quad x_2 = \begin{pmatrix} 1.58990 \\ 1.27974 \\ 0.28719 \end{pmatrix} .$$

To illustrate the scheme for minimizing $S(\theta,\hat{\Sigma})$ using a univariate nonlinear program, recall that

$$\hat{P} = \begin{pmatrix} 3.7658 & -3.8585 \\ 0 & 3.3716 \end{pmatrix} \qquad \text{(from Figure 1c)}$$

whence

$$"y_1" = (3.7648,\ -3.8585)\begin{pmatrix}2.45829\\1.59783\end{pmatrix} = 3.08980$$

$$"y_2" = (\quad 0,\ 3.3716)\begin{pmatrix}2.45829\\1.59783\end{pmatrix} = 5.38733$$

$$"y_3" = (3.7648,\ -3.8585)\begin{pmatrix}1.82933\\0.89091\end{pmatrix} = 3.44956$$

$$"y_4" = (\quad 0,\ 3.3716)\begin{pmatrix}1.82933\\0.89091\end{pmatrix} = 3.00382$$

$$"x_1" = (3.7648,\ -3.8585,\ 2.11747,\ 1.80731,\ 0.81476)'$$

$$"x_2" = (\quad 0,\ 3.3716,\ 2.11747,\ 1.80731,\ 0.81476)'$$

$$"x_3" = (3.7648,\ -3.8585,\ 1.58990,\ 1.27974,\ 0.28719)'$$

$$"x_4" = (\quad 0,\ 3.3716,\ 1.58990,\ 1.27974,\ 0.28719)'$$

$$"f"("x_1",\theta) = (3.7648)\ \ell n[(a_1 + x_1'b_{(1)})/(-1 + x_1'b_{1(3)})]$$

$$-(3.8585)\ \ell n[(a_2 + x_1'b_{(2)})/(-1 + x_1'b_{2(3)})]$$

$$"f"("x_2",\theta) = (3.3716)\ \ell n[(a_2 + x_1'b_{(2)})/(-1 + x_1'b_{2(3)})] \ .$$

SAS code to implement this scheme is shown in Figure 2a together with the resulting output.

Least squares methods lean rather heavily on normality for their validity. Accordingly, it is a sensible precaution to check residuals for evidence of severe departures from normality. Figure 2a includes a residual analysis of the unconstrained fit. There does not appear to be a gross departure from normality. Notably, the Kolmogorov-Smirnov test does not reject normality.

Figure 2a. Example 1 Fitted by Multivariate Least Squares, Unconstrained.

SAS Statements:

```
DATA WORK01;  SET EXAMPLE1;
P1=3.764814163903;  P2=-3.85846955764;  Y=P1*Y1+P2*Y2;  OUTPUT;
P1=0;  P2=3.371649857133;  Y=P1*Y1+P2*Y2;  OUTPUT;  DELETE;
PROC NLIN DATA=WORK01 METHOD=GAUSS ITER=50 CONVERGENCE=1.E-8;
PARMS B11=-.8 B12=-1.4 B13=2  B131=.5 B132=-.2 B133=-.4
      B21=.4  B22=-1.3 B23=.7 B231=.2 B232=.1  B233=-.4
      A1=-2   A2=-1;   A3=-1;
PEAK =A1+B11*X1+B12*X2+B13*X3;  BASE1=A3+B131*X1+B132*X2+B133*X3;
INTER=A2+B21*X1+B22*X2+B23*X3;  BASE2=A3+B231*X1+B232*X2+B233*X3;
MODEL Y=P1*LOG(PEAK/BASE1)+P2*LOG(INTER/BASE2);
DER.A1 =P1/PEAK;         DER.A2 =P2/INTER;
DER.B11=P1/PEAK*X1;      DER.B21=P2/INTER*X1;
DER.B12=P1/PEAK*X2;      DER.B22=P2/INTER*X2;
DER.B13=P1/PEAK*X3;      DER.B23=P2/INTER*X3;
DER.B131=-P1/BASE1*X1;  DER.B231=-P2/BASE2*X1;
DER.B132=-P1/BASE1*X2;  DER.B232=-P2/BASE2*X2;
DER.B133=-P1/BASE1*X3;  DER.B233=-P2/BASE2*X3;
OUTPUT OUT=WORK02 RESIDUAL=EHAT;
PROC UNIVARIATE DATA=WORK02 PLOT NORMAL;  VAR EHAT;  ID T;
```

Output:

```
          S T A T I S T I C A L   A N A L Y S I S   S Y S T E M          1

                NON-LINEAR LEAST SQUARES ITERATIVE PHASE

          DEPENDENT VARIABLE: Y          METHOD: GAUSS-NEWTON

  ITERATION          B11          B12           B13        RESIDUAL SS
                    B131         B132          B133
                     B21          B22           B23
                    B231         B232          B233
                      A1           A2

      0      -0.80000000  -1.40000000    2.00000000     631.16222217
              0.50000000  -0.20000000   -0.40000000
              0.40000000  -1.30000000    0.70000000
              0.20000000   0.10000000   -0.40000000
             -2.00000000  -1.00000000

      .
      .
      .

      6      -2.98669756   0.90158533    1.66353998     442.65919896
              0.26718356   0.07113302   -0.47013242
              0.20848925  -1.33081849    0.85048354
              0.18931302   0.10756268   -0.40539911
             -1.52573841  -0.96432128

NOTE: CONVERGENCE CRITERION MET.
```

Figure 2a. (Continued).

NON-LINEAR LEAST SQUARES SUMMARY STATISTICS   DEPENDENT VARIABLE Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE
REGRESSION	14	6540.63880955	467.18848640
RESIDUAL	434	442.65919896	1.01995207
UNCORRECTED TOTAL	448	6983.29800851	
(CORRECTED TOTAL)	447	871.79801949	

PARAMETER	ESTIMATE	ASYMPTOTIC STD. ERROR	ASYMPTOTIC 95 % CONFIDENCE INTERVAL LOWER	UPPER
B11	-2.98669756	1.27777798	-5.49813789	-0.47525724
B12	0.90158533	1.41306196	-1.87575225	3.67892291
B13	1.66353998	1.31692369	-0.92484026	4.25192022
B131	0.26718356	0.10864198	0.05365048	0.48071663
B132	0.07113302	0.17067332	-0.26432109	0.40658712
B133	-0.47013242	0.07443325	-0.61642910	-0.32383574
B21	0.20848925	0.41968687	-0.61639469	1.03337319
B22	-1.33081849	0.48055515	-2.27533750	-0.38629949
B23	0.85048354	0.54542139	-0.22152841	1.92249550
B231	0.18931302	0.12899074	-0.06421501	0.44284105
B232	0.10756268	0.14251811	-0.17255306	0.38767842
B233	-0.40539911	0.07932153	-0.56130357	-0.24949465
A1	-1.52573841	0.98851033	-3.46863048	0.41715366
A2	-0.96432128	0.34907493	-1.65041924	-0.27822332

Figure 2a. (Continued).

```
                 S T A T I S T I C A L   A N A L Y S I S   S Y S T E M              4

                                     UNIVARIATE

VARIABLE=EHAT

                MOMENTS                                        QUANTILES(DEF=4)

N                 448  SUM WGTS          448  100% MAX   4.62368      99%      2.68474
MEAN       -1.213E-09  SUM        -5.434E-07   75% Q3   0.648524      95%      1.63032
STD DEV      0.995133  VARIANCE     0.990289   50% MED -0.057296      90%      1.18026
SKEWNESS     0.378159  KURTOSIS      1.30039   25% Q1  -0.665501      10%      -1.1508
USS           442.659  CSS           442.659    0% MIN  -3.31487       5%     -1.60095
CV         -8.205E+10  STD MEAN    0.0470156                           1%     -2.15196
T:MEAN=0   -2.580E-08  PROB>|T|            1  RANGE      7.93855
SGN RANK        -1017  PROB>|S|     0.710841  Q3-Q1      1.31402
NUM ^= 0          448                         MODE      -3.31487
D:NORMAL    0.0310503  PROB>D          >0.15

                                      EXTREMES

                     LOWEST     ID                  HIGHEST     ID
                   -3.31487(      181)             2.65142(       152)
                   -2.44543(      128)             2.71678(        60)
                   -2.31033(       13)              2.7925(        81)
                   -2.21378(      183)             3.96674(       164)
                   -2.08762(       13)             4.62368(       132)

               BAR CHART                                              #      BOXPLOT
      4.75+*                                                          1         *
          .*                                                          1
          .                                                                     0
          .
          .**                                                         4         0
          .***                                                        5         |
          .*********                                                 17         |
          .********************                                      39         |
      0.75+****************************************                  80     +-----+
          .********************************                          63     |     |
          .***********************************************           93     *--+--*
          .*************************************                     74     +-----+
          .*********************                                     42        |
          .*********                                                 18        |
          .*****                                                     10        |
          .
     -3.25+*                                                          1         0
           ----+----+----+----+----+----+----+----+----+--
           * MAY REPRESENT UP TO 2 COUNTS

VARIABLE=EHAT

                                NORMAL PROBABILITY PLOT
          4.75+                                                        *
              |
              |                                                        *
              |
              |                                                     ****
              |                                                   **++++
              |                                               ****+
              |                                          ******
          0.75+                                   *******
              |                               ****+
              |                         *******
              |                   *******
              |              *****+
              |         *****+
              |****+
              |
         -3.25+*
               +----+----+----+----+----+----+----+----+----+----+
                   -2        -1        +0        +1        +2
```

Consider, now, fitting the model subject to the restriction that $b_{(3)}$ is the same in both equations, viz

$$H_1: b_{131} = b_{231},\ b_{132} = b_{232},\ b_{133} = b_{233} \quad .$$

As we have seen before, there are two approaches. The first is to leave the model as written and impose the constraint using the functional dependency

$$\theta = \begin{pmatrix} a_1 \\ b_{11} \\ b_{12} \\ b_{13} \\ b_{131} \\ b_{132} \\ b_{133} \\ a_2 \\ b_{21} \\ b_{22} \\ b_{23} \\ b_{231} \\ b_{232} \\ b_{233} \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \rho_4 \\ \rho_5 \\ \rho_6 \\ \rho_7 \\ \rho_8 \\ \rho_9 \\ \rho_{10} \\ \rho_{11} \\ \rho_5 \\ \rho_6 \\ \rho_7 \end{pmatrix} = g(\rho) \quad .$$

One fits the model

$$y_t = f[x_t, g(\rho^\circ)] + e_t$$

by minimizing $S[g(\rho), \hat{\Sigma}]$; derivatives are computed using the chain rule

$$\begin{aligned}
&(\partial/\partial\rho)\text{"}f\text{"}[\text{"}x_s\text{"}, g(\rho)] \\
&\quad = (\partial/\partial\rho)\, P'_{(\alpha)} f[x_t, g(\rho)] \\
&\quad = P'_{(\alpha)} (\partial/\partial\theta') f(x_t, \theta)\Big|_{\theta=g(\rho)} (\partial/\partial\rho') g(\rho) \quad .
\end{aligned}$$

These ideas were illustrated in Figure 9b of Chapter 1 and will be seen again in Figure 2d below.

The second approach is to simply rewrite the model with the constraint imposed. We adopt the second alternative, viz.

$$f(x_t,\theta) = \begin{pmatrix} \ln \dfrac{a_1 + b_{11}x_{1t} + b_{12}x_{2t} + b_{13}x_{3t}}{-1 + b_{31}x_{1t} + b_{32}x_{2t} + b_{33}x_{3t}} \\ \\ \ln \dfrac{a_2 + b_{21}x_{1t} + b_{22}x_{2t} + b_{23}x_{3t}}{-1 + b_{31}x_{1t} + b_{32}x_{2t} + b_{33}x_{3t}} \end{pmatrix}$$

$$\theta' = (a_1, b_{11}, b_{12}, b_{13}, a_2, b_{21}, b_{22}, b_{23}, b_{31}, b_{32}, b_{33}) \ .$$

SAS code to fit this model is shown in Figure 2b.

Following these, same ideas we impose the additional constraint of symmetry,

$$H_2\colon b_{12} = b_{21},\ b_{13} = b_{31},\ b_{23} = b_{32}$$

by writing

$$f(x_t,\theta) = \begin{pmatrix} \ln \dfrac{a_1 + b_{11}x_{1t} + b_{12}x_{2t} + b_{13}x_{3t}}{-1 + b_{13}x_{1t} + b_{23}x_{2t} + b_{33}x_{3t}} \\ \\ \ln \dfrac{a_2 + b_{12}x_{1t} + b_{22}x_{2t} + b_{23}x_{3t}}{-1 + b_{13}x_{1t} + b_{23}x_{2t} + b_{33}x_{3t}} \end{pmatrix}$$

$$\theta' = (a_1, b_{11}, b_{12}, b_{13}, a_2, b_{22}, b_{23}, b_{33}) \ .$$

SAS code is shown in Figure 2c.

The last restriction to be imposed, in addition to H_1 and H_2, is the homogeneity restriction

$$\Sigma_{i=1}^{3} a_i = -1 \ , \quad \Sigma_{j=1}^{3} b_{ij} = 0 \qquad \text{for } i = 1, 2, 3 \ .$$

Figure 2b. Example 1 Fitted by Multivariate Least Squares, H1 Imposed.

SAS Statements:

```
DATA WORK01;  SET EXAMPLE1;
P1=3.764814163903;  P2=-3.85846955764;  Y=P1*Y1+P2*Y2;  OUTPUT;
P1=0;  P2=3.371649857133;  Y=P1*Y1+P2*Y2;  OUTPUT;  DELETE;
PROC NLIN DATA=WORK01 METHOD=GAUSS ITER=50 CONVERGENCE=1.E-8;
PARMS B11=-.8 B12=-1.4 B13=2 B21=.4 B22=-1.3 B23=.7 B31=.5 B32=-.2 B33=-.4
      A1=-2 A2=-1;  A3=-1;
PEAK=A1+B11*X1+B12*X2+B13*X3;  INTER=A2+B21*X1+B22*X2+B23*X3;
BASE=A3+B31*X1+B32*X2+B33*X3;
MODEL Y=P1*LOG(PEAK/BASE)+P2*LOG(INTER/BASE);
DER.A1 =P1/PEAK;     DER.A2 =P2/INTER;
DER.B11=P1/PEAK*X1;  DER.B21=P2/INTER*X1;  DER.B31=(-P1-P2)/BASE*X1;
DER.B12=P1/PEAK*X2;  DER.B22=P2/INTER*X2;  DER.B32=(-P1-P2)/BASE*X2;
DER.B13=P1/PEAK*X3;  DER.B23=P2/INTER*X3;  DER.B33=(-P1-P2)/BASE*X3;
```

Output:

```
            S T A T I S T I C A L   A N A L Y S I S   S Y S T E M            1

                  NON-LINEAR LEAST SQUARES ITERATIVE PHASE

              DEPENDENT VARIABLE: Y          METHOD: GAUSS-NEWTON

  ITERATION          B11            B12            B13          RESIDUAL SS
                     B21            B22            B23
                     B31            B32            B33
                      A1             A2

      0          -0.80000000    -1.40000000     2.00000000     641.48045300
                  0.40000000    -1.30000000     0.70000000
                  0.50000000    -0.20000000    -0.40000000
                 -2.00000000    -1.00000000

      .
      .
      .

      8          -3.27643190     1.30488351     1.66561680     447.31829119
                  0.40180449    -1.11931853     0.41058766
                  0.23944183     0.10626154    -0.45982238
                 -1.58236942    -1.20266408

NOTE: CONVERGENCE CRITERION MET.
            S T A T I S T I C A L   A N A L Y S I S   S Y S T E M            2

    NON-LINEAR LEAST SQUARES SUMMARY STATISTICS     DEPENDENT VARIABLE Y

     SOURCE                  DF       SUM OF SQUARES        MEAN SQUARE

     REGRESSION              11       6535.97971731        594.17997430
     RESIDUAL               437        447.31829119          1.02361165
     UNCORRECTED TOTAL      448       6983.29800851

     (CORRECTED TOTAL)      447        871.79801949

PARAMETER          ESTIMATE        ASYMPTOTIC              ASYMPTOTIC 95 %
                                   STD. ERROR           CONFIDENCE INTERVAL
                                                      LOWER             UPPER
B11             -3.27643190        1.27198559      -5.77643950      -0.77642431
B12              1.30488351        0.95321400      -0.56859860       3.17836562
B13              1.66561680        1.01449051      -0.32830042       3.65953403
B21              0.40180449        0.29689462      -0.18172319       0.98533217
B22             -1.11931853        0.36601761      -1.83870310      -0.39993395
B23              0.41058766        0.33172431      -0.24139558       1.06257090
B31              0.23944183        0.09393101       0.05482635       0.42405730
B32              0.10626154        0.11605620      -0.12183961       0.33436270
B33             -0.45982238        0.05409256      -0.56613790      -0.35350687
A1              -1.58236942        0.85859333      -3.26988056       0.10514171
A2              -1.20266408        0.23172071      -1.65809655      -0.74723161
```

Figure 2c. Example 1 Fitted by Multivariate Least Squares, H1 and H2 Imposed.

SAS Statements:

```
DATA WORK01;  SET EXAMPLE1;
P1=3.764814163903;  P2=-3.85846955764;  Y=P1*Y1+P2*Y2;  OUTPUT;
P1=0;  P2=3.371649857133;  Y=P1*Y1+P2*Y2;  OUTPUT;  DELETE;
PROC NLIN DATA=WORK01 METHOD=GAUSS ITER=50 CONVERGENCE=1.E-8;
PARMS B11=0 B12=0 B13=0 B22=0 B23=0 B33=0 A1=-1 A2=-1;  A3=-1;
PEAK=A1+B11*X1+B12*X2+B13*X3;  INTER=A2+B12*X1+B22*X2+B23*X3;
BASE=A3+B13*X1+B23*X2+B33*X3;
MODEL Y=P1*LOG(PEAK/BASE)+P2*LOG(INTER/BASE);
DER.A1 =P1/PEAK;  DER.A2 =P2/INTER;  DER.B11=P1/PEAK*X1;
DER.B12=P1/PEAK*X2+P2/INTER*X1;  DER.B22=P2/INTER*X2;
DER.B13=P1/PEAK*X3+(-P1-P2)/BASE*X1;  DER.B23=P2/INTER*X3+(-P1-P2)/BASE*X2;
DER.B33=(-P1-P2)/BASE*X3;
```

Output:

```
          S T A T I S T I C A L   A N A L Y S I S   S Y S T E M          1

                NON-LINEAR LEAST SQUARES ITERATIVE PHASE

           DEPENDENT VARIABLE: Y          METHOD: GAUSS-NEWTON

   ITERATION          B11              B12              B13        RESIDUAL SS
                      B22              B23              B33
                       A1               A2

       0         0.000000E+00     0.000000E+00     0.000000E+00   6983.29800851
                 0.000000E+00     0.000000E+00     0.000000E+00
                  -1.00000000      -1.00000000

       .
       .
       .

      11          -1.28362479       0.81889299       0.36106759    450.95423403
                  -1.04835591       0.03049767      -0.46735947
                  -2.92727122      -1.53786463

NOTE: CONVERGENCE CRITERION MET.

          S T A T I S T I C A L   A N A L Y S I S   S Y S T E M          3

     NON-LINEAR LEAST SQUARES SUMMARY STATISTICS     DEPENDENT VARIABLE Y
```

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE
REGRESSION	8	6532.34377448	816.54297181
RESIDUAL	440	450.95423403	1.02489599
UNCORRECTED TOTAL	448	6983.29800851	
(CORRECTED TOTAL)	447	871.79801949	

PARAMETER	ESTIMATE	ASYMPTOTIC STD. ERROR	ASYMPTOTIC 95 % CONFIDENCE INTERVAL LOWER	UPPER
B11	-1.28362479	0.22679435	-1.72936637	-0.83788321
B12	0.81889299	0.08096691	0.65976063	0.97802535
B13	0.36106759	0.03024703	0.30162008	0.42051510
B22	-1.04835591	0.08359301	-1.21264961	-0.88406221
B23	0.03049767	0.03608943	-0.04043249	0.10142783
B33	-0.46735947	0.01923198	-0.50515801	-0.42956093
A1	-2.92727122	0.27778075	-3.47322147	-2.38132098
A2	-1.53786463	0.09167461	-1.71804189	-1.35768737

As we have noted, the scaling convention is irrelevant as far as the data is concerned. The restriction $a_1 + a_2 + a_3 = -1$ is just a scaling convention and, other than asthetics, there is no reason to prefer it to the convention $a_3 = -1$ that we have imposed thus far.* Retaining $a_3 = -1$, the hypothesis of homogeneity can be rewritten as the parametric restriction

$$H_3: \Sigma_{i=1}^{3} b_{ij} = 0 \quad \text{for } j = 1, 2, 3 \quad .$$

Equivalently, H_3 can be written as the functional dependency

$$\theta = \begin{pmatrix} a_1 \\ b_{11} \\ b_{12} \\ b_{13} \\ a_2 \\ b_{22} \\ b_{23} \\ b_{33} \end{pmatrix} = \begin{pmatrix} a_1 \\ -b_{12} - b_{13} \\ b_{12} \\ b_{13} \\ a_2 \\ -b_{23} - b_{12} \\ b_{23} \\ -b_{23} - b_{13} \end{pmatrix} = \begin{pmatrix} \rho_1 \\ -\rho_2 - \rho_3 \\ \rho_2 \\ \rho_3 \\ \rho_4 \\ -\rho_5 - \rho_2 \\ \rho_5 \\ -\rho_5 - \rho_3 \end{pmatrix} = g(\rho)$$

with Jacobian

$$G(\rho) = (\partial/\partial\rho')g(\rho) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 & -1 \end{pmatrix} .$$

SAS code implementing this restriction is shown in Figure 2d.

---

*In economic parlance, it is impossible to tell the difference between a linear homogeneous and a homothetic indirect utility function by looking at a demand system.

Figure 2d. Example 1 Fitted by Multivariate Least Squares, H1, H2, and H3 Imposed.

SAS Statements:

```
DATA WORK01;  SET EXAMPLE1;
P1=3.764814163903;  P2=-3.85846955764;  Y=P1*Y1+P2*Y2;  OUTPUT;
P1=0;  P2=3.371649857133;  Y=P1*Y1+P2*Y2;  OUTPUT;  DELETE;
PROC NLIN DATA=WORK01 METHOD=GAUSS ITER=50 CONVERGENCE=1.E-8;
PARMS R1=-3 R2=.8 R3=.4 R4=-1.5 R5=.03;  A3=-1;
A1=R1; B11=-R2-R3; B12=R2; B13=R3; A2=R4; B22=-R5-R2; B23=R5; B33=-R5-R3;
PEAK=A1+B11*X1+B12*X2+B13*X3;  INTER=A2+B12*X1+B22*X2+B23*X3;
BASE=A3+B13*X1+B23*X2+B33*X3;
MODEL Y=P1*LOG(PEAK/BASE)+P2*LOG(INTER/BASE);
DER_A1 =P1/PEAK;  DER_A2 =P2/INTER;  DER_B11=P1/PEAK*X1;
DER_B12=P1/PEAK*X2+P2/INTER*X1;  DER_B22=P2/INTER*X2;
DER_B13=P1/PEAK*X3+(-P1-P2)/BASE*X1;  DER_B23=P2/INTER*X3+(-P1-P2)/BASE*X2;
DER_B33=(-P1-P2)/BASE*X3;
DER.R1=DER_A1; DER.R2=-DER_B11+DER_B12-DER_B22; DER.R3=-DER_B11+DER_B13-DER_B33;
DER.R4=DER_A2; DER.R5=-DER_B22+DER_B23-DER_B33;
```

Output:

S T A T I S T I C A L   A N A L Y S I S   S Y S T E M   1

NON-LINEAR LEAST SQUARES ITERATIVE PHASE

DEPENDENT VARIABLE: Y   METHOD: GAUSS-NEWTON

ITERATION	R1 R4	R2 R5	R3	RESIDUAL SS
0	-3.00000000 -1.50000000	0.80000000 0.03000000	0.40000000	560.95959664
1	-2.70479542 -1.59049135	0.85805995 0.05440110	0.37705186	478.82185398
2	-2.72429979 -1.59215078	0.85764215 0.05770545	0.37433263	478.79661265
3	-2.72517893 -1.59211417	0.85757097 0.05794017	0.37413074	478.79654696
4	-2.72523507 -1.59210976	0.85756494 0.05795637	0.37411703	478.79654666

NOTE: CONVERGENCE CRITERION MET.

S T A T I S T I C A L   A N A L Y S I S   S Y S T E M   2

NON-LINEAR LEAST SQUARES SUMMARY STATISTICS   DEPENDENT VARIABLE Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE
REGRESSION	5	6504.50146185	1300.90029237
RESIDUAL	443	478.79654666	1.08080485
UNCORRECTED TOTAL	448	6983.29800851	
(CORRECTED TOTAL)	447	871.79801949	

PARAMETER	ESTIMATE	ASYMPTOTIC STD. ERROR	ASYMPTOTIC 95 % CONFIDENCE INTERVAL LOWER	UPPER
R1	-2.72523507	0.17799072	-3.07505148	-2.37541867
R2	0.85756494	0.06718212	0.72552768	0.98960220
R3	0.37411703	0.02709873	0.32085818	0.42737587
R4	-1.59210976	0.07719388	-1.74382378	-1.44039575
R5	0.05795637	0.03403316	-0.00893116	0.12484390

Table 2. Parameter Estimates and Standard Errors for Example 1.

Parameter	Unconstrained	Subject to: H1	H1 & H2	H1, H2 & H3
A1	-1.5257 (0.9885)	-1.5824 (0.8586)	-2.9273 (0.2778)	-2.7252 (0.1780)
B11	-2.9867 (1.2778)	-3.2764 (1.2720)	-1.2836 (0.2268)	-1.2317
B12	0.9016 (1.4131)	1.3049 (0.9532)	0.8189 (0.0810)	0.8576 (0.0672)
B13	1.6635 (1.3169)	1.6656 (1.0145)	0.3611 (0.0302)	0.3741 (0.0271)
A2	-0.9643 (0.3491)	-1.2027 (0.2317)	-1.5379 (0.0917)	-1.5921 (0.0772)
B21	0.2085 (0.4197)	0.4018 (0.2969)	0.8189	0.8576
B22	-1.3308 (0.4806)	-1.1193 (0.3660)	-1.0484 (0.0836)	-0.9155
B23	0.8505 (0.5454)	0.4106 (0.3317)	0.0305 (0.0361)	0.0580 (0.0340)
A3	-1.0 (0.0)	-1.0 (0.0)	-1.0 (0.0)	-1.0 (0.0)
B131/B31	0.2672 (0.1086)	0.2394 (0.0939)	0.3611	0.3741
B132/B32	0.0711 (0.1707)	0.1063 (0.1161)	0.0305	0.0580
B133/B33	-0.4701 (0.0744)	-0.4598 (0.0541)	-0.4674 (0.0192)	-0.4321
B231	0.1893 (0.1290)			
B232	0.1076 (0.1425)			
B233	-0.4054 (0.0793)			

Standard errors shown in parentheses.

As was seen in Chapter 1, these regressions can be assessed using the likelihood ratio test statistic

$$L = \frac{(SSE_{reduced} - SSE_{full})/q}{(SSE_{full})/(n-p)} \quad .$$

As with linear regression, when one has a number of such tests to perform it is best to organize them into an analysis of variance table as shown in Table 3. For each hypothesis listed under source in Table 3, the entry listed under d. f. is q , as above, and that listed under Sum of Squares is $(SSE_{reduced} - SSE_{full})$, as above. As an instance, to test H1, H2 & H3 jointly one has

$$SSE_{reduced} = 478.79654666 \quad \text{(from Figure 2d)}$$
$$SSE_{full} = 442.65919896 \quad \text{(from Figure 2a)}$$

with 443 and 434 degrees of freedom respectively which yields

$$(SSE_{reduced} - SSE_{full}) = 36.1374$$
$$q = 443 - 434 = 9$$

as shown in Table 3. In general, the mean sum of squares cannot be split from the total regression sum of squares but in this instance it would be possible to fit a mean to the data as a special case of the nonlinear model by setting B = 0 and choosing

$$a = {}_{3}\begin{pmatrix} -\hat{P}^{-1}\begin{pmatrix} e^{\mu} \\ e^{\mu} \end{pmatrix} \\ -1 \end{pmatrix}_{1} \quad .$$

The existence of a parametric restriction that will produce the model "y_s" $= \mu\binom{1}{1} +$"e_s" justifies the split. The sum of squares for the mean is computed from

Table 3. Analysis of Variance

Source	d.f.	Sum of Squares	Mean Square	F	P>F
Mean	1	6111.5000	6111.5000		
Regression	4	393.0015	98.2504	96.324	0.00000
H1, H2, & H3	9	36.1374	4.0153	3.937	0.0001
H1	3	4.6591	1.5530	1.523	0.206
H2 after H1	3	3.6360	1.2120	1.188	0.313
H3 after H1 & H2	3	27.8423	9.2808	9.099	0.00001
Error	434	442.6592	1.0200		
Total	448	6983.2980			

$$SSE_{reduced} = 6983.29800851 \quad \text{(from Figure 2d)}$$

$$SSE_{full} = 871.798001949 \quad \text{(from Figure 2d)}$$

with 448 and 447 degrees of freedom respectively yielding

$$(SSE_{reduced} - SSE_{full}) = 6111.5000$$

$$q = 448 - 447 = 1$$

which is subtracted from

$$SSE_{regression} = 6504.50146185 \quad \text{(from Figure 2d)}$$

with 5 degrees of freedom to yield the entry shown in Table 3.

From Table 3 one sees that the model of Figure 2c is reasonably well supported by the data and the model of Figure 2d is not. Accordingly, we shall accept it as adequate throughout most of the rest of the sequel realizing that there are potential specification errors of at least two sorts. The first are omitted variables of which those listed in Table 1c are prime candidates and the second is an erroneous specification of functional form. But our purpose is illustrative and we shall not dwell on this matter. The model of Figure 2c will serve. □

As suggested by the preceding analysis, in the sequel we shall accept the information provided by $\hat{\Sigma}^{-1} = \hat{P}'\hat{P}$ regarding the rotation $\hat{P}$ that will reduce the multivariate model to a univariate model, as we must to make any progress, but we shall disregard the scale information and shall handle scaling in accordance with standard practice for univariate models. To state this differently, in using Table 3 we could have entered a table of the chi-square distribution using 27.8423 with 3 degrees of freedom but instead we entered a table of the F-distribution using 9.099 with 3 and 434 degrees of freedom.

The idea of rewriting the multivariate model

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

in the form

$$"y_s" = "f"("x_s",\theta) + "e_s" \qquad s = 1, 2, \ldots, nM$$

using the transformation

$$"y_s" = p'_{(\alpha)} y_t \qquad s = M(t-1) + \alpha$$

in order to be able to use univariate nonlinear regression methods is useful pedagogically and is even convenient for small values of M . In general, however, one needs to be able to minimize $S(\theta,\Sigma)$ directly. To do this note that the Gauss-Newton correction vector is, from Section 4 of Chapter 1,

$$\begin{aligned}
D(\theta,\Sigma) &= \{\Sigma_{s=1}^{nM}[(\partial/\partial\theta)\ "f"("x_s",\theta)][(\partial/\partial\theta)\ "f"("x_s",\theta)]'\}^{-1} \\
&\quad \times \Sigma_{s=1}^{nM}[(\partial/\partial\theta)\ "f"("x_s",\theta)]["y_s" - "f"("x_s",\theta)] \\
&= \{\Sigma_{t=1}^{n}\ \Sigma_{\alpha=1}^{M}[(\partial/\partial\theta)f'(x_t,\theta)]p_{(\alpha)}p'_{(\alpha)}[(\partial/\partial\theta')f(x_t,\theta)]\}^{-1} \\
&\quad \times \Sigma_{t=1}^{n}\ \Sigma_{\alpha=1}^{M}[(\partial/\partial\theta')f(x_t,\theta)]p_{(\alpha)}p'_{(\alpha)}[y_t - f(x_t,\theta)] \\
&= \{\Sigma_{t=1}^{n}[(\partial/\partial\theta)f'(x_t,\theta)]\ \Sigma^{-1}[(\partial/\partial\theta')f(x_t,\theta)]\}^{-1} \\
&\quad \times \Sigma_{t=1}^{n}[(\partial/\partial\theta)f'(x_t,\theta)]\ \Sigma^{-1}[y_t - f(x_t,\theta)] \ .
\end{aligned}$$

The modified Gauss-Newton algorithm for minimizing $S(\theta,\Sigma)$ is, then:

0) Choose a starting estimate θ_0 . Compute $D_0 = D(\theta_0,\Sigma)$ and find a λ_0 between zero and one such that

$$S(\theta_0 + \lambda_0 D_0,\Sigma) < S(\theta_0,\Sigma) \ .$$

1) Let $\theta_1 = \theta_0 + \lambda_0 D_0$. Compute $D_1 = D(\theta_1, \Sigma)$ and find a λ_1 between zero and one such that

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

2) Let $\theta_2 = \theta_1 D_1$.

.
.
.

The comments in Section 4 of Chapter 1 regarding starting rules, stopping rules, alternative algorithms apply directly.

1) Let $\theta_1 = \theta_0 + \lambda_0 D_0$. Compute $D_1 = D(\theta_1,\Sigma)$ and find a λ_1 between zero and one such that

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

2) Let $\theta_2 = \theta_1 D_1$.

.
.
.

The comments in Section 4 of Chapter 1 regarding starting rules, stopping rules, alternative algorithms apply directly.

PROBLEMS

1. Show that if e_α is an n-vector with typical element $e_{\alpha t}$ for $t = 1, 2, \ldots, n$ and $\alpha = 1, 2, \ldots, M$ and $C(e_{\alpha t}, e_{\beta s}) = \sigma_{\alpha\beta}$ if $t = s$ and is zero otherwise then

$$C(e_\alpha, e'_\beta) = \sigma_{\alpha\beta}\, {}_nI_n \quad .$$

2. Re-estimate Example 1 (in unconstrained form, subject to H_1, subject to H_1 & H_2, and subject to H_1, H_2, & H_3) using the normalizing convention $a_1 + a_2 + a_3 = -1$ (instead of $a_3 = -1$ as used in Figures 2a, 2b, 2c, 2d).

3. Using the "seemingly unrelated" regressions notation, show that the Gauss-Newton correction vector can be written as

$$D(\theta,\Sigma) = [F'(\theta)(\Sigma \otimes I)\, F(\theta)]^{-1}\, F'(\theta)(\Sigma \otimes I)[y - f(\theta)]$$

where $F(\theta) = (\partial/\partial\theta')f(\theta) = \text{diag}\,[F_1(\theta_1), \ldots, F_M(\theta_M)]$.

4. Show that $S(\theta,\Sigma)$ satisfies Assumption 4 of Chapter 3 which justifies the expedient of replacing Σ by $\hat{\Sigma}$ and subsequently acting as if $\hat{\Sigma}$ were the true value of Σ .

5. If the model used in Example 1 is misspecified as to choice of functional form then theory suggests (Gallant, 1981, 1982; Elbadawi, Gallant, and Souza, 1983) that the misspecification must take the form of omission of additive terms of the form

$$a_{\alpha j} \cos(jk'_\alpha x) - b_{\alpha j} \sin(jk'_\alpha x)$$

from the indirect utility function

$$g(x|\theta) = a'x + (\tfrac{1}{2})x'B\,x\ ;$$

recall that $x = \ell n(p/E)$. Test for the joint omission of these terms for

$$k_\alpha = \begin{pmatrix}1\\0\\0\end{pmatrix},\ \begin{pmatrix}0\\1\\0\end{pmatrix},\ \begin{pmatrix}0\\0\\1\end{pmatrix},\ \begin{pmatrix}1\\1\\0\end{pmatrix},\ \begin{pmatrix}1\\0\\1\end{pmatrix},\ \begin{pmatrix}0\\1\\1\end{pmatrix}$$

and j = 1, 2, a total of 24 additional parameters.

6. Instead of splitting out one degree of freedom for the model

$$"y_s" = \mu(\tfrac{1}{1}) + "e_s"$$

from the five degrees of freedom regression sum of squares of Figure 2d as was done in Table 3, split out two degrees of freedom for the model

$$"y_s" = \begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix} + "e_s" \quad .$$

7. (Out of range argument) Show that the constants t_1, t_2, a, b, c, α, β can be chosen so that the function

$$\text{slog}(x) = \begin{cases} \alpha + \beta x & -\infty < x \leqq t_1 \\ a + bx + cx^2 & t_1 \leqq x \leqq t_2 \\ \ell n\ x & t_2 \leqq x < \infty \end{cases}$$

is continuous with continuous first derivative

$$(d/dx)\text{slog}(x) = \begin{cases} \beta & -\infty < x \leqq t_1 \\ b + 2cx & t_1 \leqq x \leqq t_2 \\ 1/x & t_2 \leqq x < \infty \end{cases} \quad .$$

Verify that slog(x) is once continuously differentiable if the constants are chosen as

```
T2 = 1.D-7
T1 = 0.D0
ALPHA=-299.999999999999886D0
BETA=5667638086.9808321D0
A=-299.999999999999886D0
B=5667638086.9808321D0
C=-28288190434904165.D0
```

Use slog(x) in place of $\ell n(x)$ in Figures 1a through 2d and observe that the same numerical results obtain.

3. ASYMPTOTIC THEORY

As in Chapter 4, an asymptotic estimation theory obtains by restating Assumptions 1 through 6 of Chapter 3 in context and then applying Theorems 3 and 5 of Chapter 3. Similarly, one obtains an asymptotic theory of inference by appending restatements of Assumptions 7 and 13 to the list of assumptions and then applying Theorems 11, 14, and 15.

Of the two notational schemes, "seemingly unrelated" and multivariate, the multivariate is the more convenient to this task. Recall, that in this scheme the model is written as

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

with θ° known to lie in some compact set Θ^*. The functional form of $f(x,\theta)$ is known, x is k-dimensional, θ is p-dimensional, and $f(x,\theta)$ takes its values in R^M; y_t and e_t are M-vectors. The errors e_t are independently and identically distributed each with mean zero and nonsingular variance-covariance matrix Σ; viz.

$$\mathcal{E} e_t = 0, \quad \mathcal{C}(e_t, e_s') = \begin{cases} \Sigma & t = s \\ 0 & t \neq s \end{cases} .$$

The parameter θ° is estimated by $\hat{\theta}_n$ that minimizes

$$s_n(\theta, \hat{\Sigma}_n) = (1/n)\Sigma_{t=1}^n [y_t - f(x_t,\theta)]' \hat{\Sigma}_n^{-1} [y_t - f(x_t,\theta)] .$$

Here we shall let $\hat{\Sigma}_n$ be any random variable that converges almost surely to Σ and has $\sqrt{n}\,(\hat{\Sigma}_n - \Sigma)$ bounded in probability; that is, given $\delta > 0$ there is a bound b and a sample size N such that

$$P(\sqrt{n}\,|\hat{\sigma}_{\alpha\beta n} - \sigma_{\alpha\beta}| < b) \geq 1 - \delta$$

for all $n > N$, $\sigma_{\alpha\beta}$ being a typical element of Σ. Verification that the estimator of Σ proposed in the previous section satisfies this requirement is left to Problem 1.

Construction of the set T which is presumed to contain $\hat{\Sigma}_n$ requires a little care. Denote the upper triangle of Σ by

$$\tau = (\sigma_{11},\sigma_{12},\sigma_{22},\sigma_{13},\sigma_{23},\sigma_{33},\ldots,\sigma_{1M},\sigma_{2M},\ldots,\sigma_{MM})'$$

which is a column vector of length $M(M+1)/2$. Let $\Sigma(\tau)$ denote the mapping of τ into the elements of Σ and set $\Sigma(\hat{\tau}_n) = \hat{\Sigma}_n$, $\Sigma(\tau^*) = \mathcal{C}(e_t,e_t')$. Now $\det \Sigma(\tau)$ is a polynomial of degree M in τ and is therefore continuous; moreover for some $\delta > 0$ we have $\det \Sigma(\tau^*) - \delta > 0$ by assumption. Therefore the set

$$\{\tau : \det \Sigma(\tau) > \det \Sigma(\tau^*) - \delta\}$$

is an open set containing τ^*. Then this set must contain a bounded open ball with center τ^* and the closure of this ball can be taken as T. The assumption that $\sqrt{n}\,(\hat{\Sigma}_n - \Sigma)$ is bounded in probability means that we have implicitly taken $\tau_n^o \equiv \tau^* \in T$ and, without loss of generality (Problem 2), we can assume that $\hat{\tau}_n$ is in T for all n. Note that $\det \Sigma(\tau) \ge \det \Sigma(\tau^*) - \delta$ for all τ in T which implies that $\Sigma^{-1}(\tau)$ is continuous and differentiable over T (Problem 3). Put

$$B = \sup\{\sigma^{\alpha\beta}(\tau) : \tau \in T,\ \alpha,\beta = 1,2,\ldots,M\}$$

where $\sigma^{\alpha\beta}(\tau)$ denotes a typical element of $\Sigma^{-1}(\tau)$. Since $\Sigma^{-1}(\tau)$ is continuous over the compact set T we must have $B < \infty$.

We are interested in testing the hypothesis

$$H\colon h(\theta^o) = 0 \quad \text{against} \quad A\colon h(\theta^o) \ne 0$$

which we assume can be given the equivalent representation

$$H\colon \theta^o = g(\rho^o) \text{ for some } \rho^o \quad \text{against} \quad A\colon \theta^o \ne g(\rho) \text{ for any } \rho$$

where $h\colon \mathbb{R}^p \to \mathbb{R}^q$, $g\colon \mathbb{R}^r \to \mathbb{R}^p$, and $p = r + q$. The correspondence with the notation of Chapter 3 is as follows.

NOTATION 1

General (Chapter 3)	Specific (Chapter 6)
$e_t = q(y_t, x_t, \gamma_n^o)$	$e_t = y_t - f(x_t, \theta_n^o)$
$\gamma \in \Gamma^*$	$\theta \in \Theta^*$
$y = Y(e, x, \gamma)$	$y = f(x, \theta) + e$
$s(y_t, x_t, \hat{\tau}_n, \lambda)$	$[y_t - f(x_t, \theta)]' \Sigma^{-1}(\hat{\tau}_n)[y_t - f(x_t, \theta)]$
$\hat{\tau}_n \in T$	$\hat{\tau}_n \in T, \quad \tau_n^o \equiv \tau^*$
$\lambda \in \Lambda^*$	$\theta \in \Theta^*$
$s_n(\lambda) = (1/n)\Sigma_{t=1}^n s(y_t, x_t, \hat{\tau}_n, \lambda)$	$s_n(\theta) = (1/n)\Sigma_{t=1}^n [y_t - f(x_t, \theta)]' \Sigma^{-1}(\hat{\tau}_n)[y_t - f(x_t, \theta)]$
$s_n^o(\lambda) = (1/n)\Sigma_{t=1}^n \int_{\mathcal{E}} s[Y(e, x_t, \gamma_n^o), x_t, \tau_n^o, \lambda] dP(e)$	$s_n^o(\theta) = M + (1/n)\Sigma_{t=1}^n [f(x_t, \theta_n^o) - f(x_t, \theta)]' \Sigma^{-1}(\tau^*)[f(x_t, \theta_n^o) - f(x_t$
$s^*(\lambda) = \int_{\mathcal{X}} \int_{\mathcal{E}} s[Y(e, x, \gamma^*), x, \tau^*, \lambda] dP(e)\, d\mu(x)$	$s^*(\theta) = M + \int_{\mathcal{X}} [f(x, \theta^*) - f(x, \theta)]' \Sigma^{-1}(\tau^*)[f(x, \theta^*) - f(x, \theta)] d\mu(x$
$\hat{\lambda}_n$ minimizes $s_n(\lambda)$	$\hat{\theta}_n$ minimizes $s_n(\theta)$
$\tilde{\lambda}_n$ minimizes $s_n(\lambda)$ subject to $h(\lambda) = 0$	$\tilde{\theta}_n = g(\hat{\rho}_n)$ minimizes $s_n(\theta)$ subject to $h(\theta) = 0$
λ_n^o minimizes $s_n^o(\lambda)$	θ_n^o minimizes $s_n^o(\theta)$
λ_n^* minimizes $s_n^o(\lambda)$ subject to $h(\lambda) = 0$	$\theta_n^* = g(\rho_n^o)$ minimizes $s_n^o(\theta)$ subject to $h(\theta) = 0$
λ^* minimizes $s^*(\lambda)$	θ^* minimizes $s^*(\theta)$

Assumptions 1 through 6 of Chapter 3 read as follows in the present case.

ASSUMPTION 1′. The errors are independently and identically distributed with common distribution $P(e)$. □

ASSUMPTION 2′. $f(x,\theta)$ is continuous on $\mathcal{X} \times \Theta^*$ and Θ^* is compact. □

ASSUMPTION 3′. (Gallant and Holly, 1980) Almost every realization of $\{v_t\}$ with $v_t = (e_t, x_t)$ is a Cesaro sum generator with respect to the product measure

$$\nu(A) = \int_{\mathcal{X}} \int_{\mathcal{E}} I_A(e,x)\, dP(e)\, d\mu(x)$$

and dominating function $b(e,x)$. The sequence $\{x_t\}$ is a Cesaro sum generator with respect to μ and $b(x) = \int_{\mathcal{E}} b(e,x)\, dP(e)$. For each $x \in \mathcal{X}$ there is a neighborhood N_x such that $\int_{\mathcal{E}} \sup_{N_x} b(e,x)\, dP(e) < \infty$. □

ASSUMPTION 4′. (Identification) The parameter θ° is indexed by n and the sequence $\{\theta_n^\circ\}$ converges to θ^*; $\tau_n^\circ = \tau^*$, $\sqrt{n}(\hat{\tau}_n - \tau^*)$ is bounded in probability, and $\hat{\tau}_n$ converges almost surely to τ^*.

$$s^*(\theta) = M + \int_{\mathcal{X}} [f(x,\theta^*) - f(x,\theta)]' \Sigma^{-1}(\tau^*)[f(x,\theta^*) - f(x,\theta)]' d\mu(x)$$

has a unique minimum over Θ^* at θ^*. □

ASSUMPTION 5′. Θ^* is compact; $\{\hat{\tau}_n\}$, T, and B are as described in the first few paragraphs of this section. The functions

$$[e_\alpha + f_\alpha(x,\theta^\circ) - f_\alpha(x,\theta)][e_\beta + f_\beta(x,\theta^\circ) - f_\beta(x,\theta)]$$

are dominated by $b(e,x)/M^2 B$ over $\mathcal{E} \times \mathcal{X} \times \Theta^* \times \Theta^*$; $b(e,x)$ is that of Assumption 3′. □

This is enough to satisfy Assumption 5 of Chapter 3 since

$$
\begin{aligned}
&s[Y(e,x,\theta^\circ),x,\tau,\theta] \\
&\quad = [e + f(x,\theta^\circ) - f(x,\theta)]'\Sigma^{-1}(\tau)[e + f(x,\theta^\circ) - f(x,\theta)] \\
&\quad \le B\, \Sigma_\alpha \Sigma_\beta |e_\alpha + f_\alpha(x,\theta^\circ) - f_\beta(x,\theta)||e_\beta + f_\beta(x,\theta^\circ) - f_\beta(x,\theta)| \\
&\quad \le B\, M^2\, b(e,x)/M^2 B \\
&\quad = b(e,x)\,.
\end{aligned}
$$

The sample objective function is

$$s_n(\theta) = (1/n)\Sigma_{t=1}^n [y_t - f(x_t,\theta)]'\Sigma^{-1}(\hat\tau_n)[y_t - f(x_t,\theta)]\,.$$

Replacing $\hat\tau_n$ by $\tau_n^\circ \equiv \tau^*$, its expectation is

$$
\begin{aligned}
s_n^\circ(\theta) &= (1/n)\Sigma_{t=1}^n \mathcal{E}[e_t + f(x_t,\theta_n^\circ) - f(x_t,\theta)]'\Sigma^{-1}(\tau^*)[e_t + f(x_t,\theta_n^\circ) - f(x_t,\theta)] \\
&= (1/n)\Sigma_{t=1}^n \mathcal{E} e_t' \Sigma^{-1}(\tau^*) e_t \\
&\quad + (1/n)\Sigma_{t=1}^n [f(x_t,\theta_n^\circ) - f(x_t,\theta)]'\Sigma^{-1}(\tau^*)[f(x_t,\theta_n^\circ) - f(x_t,\theta)] \\
&= M + (1/n)\Sigma_{t=1}^n [f(x_t,\theta_n^\circ) - f(x_t,\theta)]'\Sigma^{-1}(\tau^*)[f(x_t,\theta_n^\circ) - f(x_t,\theta)];
\end{aligned}
$$

the last equality obtains from

$$
\begin{aligned}
&(1/n)\Sigma_{t=1}^n \mathcal{E} e_t' \Sigma^{-1}(\tau^*) e_t \\
&\quad = (1/n)\Sigma_{t=1}^n \operatorname{tr} \Sigma^{-1}(\tau^*)\, \mathcal{E}\, e_t e_t' \\
&\quad = (1/n)\Sigma_{t=1}^n \operatorname{tr} \underset{M\,M}{I} \\
&\quad = M\,.
\end{aligned}
$$

By Lemma 1 of Chapter 3, both $s_n(\theta)$ and $s_n^\circ(\theta)$ have uniform almost sure limit

$$s^*(\theta) = M + \int_{\mathcal{X}} [f(x,\theta^*) - f(x,\theta)]'\Sigma^{-1}(\tau^*)[f(x,\theta^*) - f(x,\theta)]d\mu(x)\,.$$

Note that the true value θ_n° of the unknown parameter is also a minimizer of

$s_n^\circ(\theta)$ so that the use of θ_n° to denote them both is not ambiguous. By Theorem 3 of Chapter 3 we have

$$\ell im_{n\to\infty}\, \hat{\theta}_n = \theta^* \quad \text{almost surely.}$$

Continuing, we have one last assumption in order to be able to claim asymptotic normality.

ASSUMPTION 6′. Θ^* contains a closed ball Θ centered at θ^* with finite, nonzero radius such that

$$[e_\alpha + f_\alpha(x,\theta^\circ) - f_\alpha(x,\theta)][(\partial/\partial\theta_i)f_\beta(x,\theta)]\,,$$

$$[(\partial/\partial\theta_i)f_\alpha(x,\partial)][(\partial/\partial\theta_j)f_\beta(x,\theta)]\,,$$

$$[e_\alpha + f_\alpha(x,\theta^\circ) - f_\alpha(x,\theta)][(\partial^2/\partial\theta_i\partial\theta_j)f(x,\theta)]\,, \text{ and}$$

$$\{[e_\alpha + f_\alpha(x,\theta^\circ) - f_\alpha(x,\theta)][(\partial/\partial\theta_i)f_\beta(x,\theta)]\}^2 \text{ are}$$

dominated by $b(e,x)/BM^2$ over $\mathcal{E} \times \mathcal{X} \times \Theta \times \Theta$ for $i,j = 1,2,\ldots, p$ and $\alpha,\beta = 1,2,\ldots,M$. Moreover,

$$\mathcal{J}^* = 2\int_{\mathcal{X}} [(\partial/\partial\theta')f(x,\theta^*)]'\Sigma^{-1}(\tau^*)[(\partial/\partial\theta)f(x,\theta^*)]\, d\mu(x)$$

is nonsingular. □

One can verify that this is enough to dominate

$$(\partial/\partial\theta_i)s[Y(e,x,\theta^\circ),x,\tau,\theta]$$
$$= -2[e + f(x,\theta^\circ) - f(x,\theta)]'\Sigma^{-1}(\tau)(\partial/\partial\theta_i)f(x,\theta)$$

$$(\partial^2/\partial\theta_i\partial\theta_j)s[Y(e,x,\theta^\circ),x,\tau,\theta]$$
$$= 2[(\partial/\partial\theta_i)f(x,\theta)]'\Sigma^{-1}(\tau)[(\partial/\partial\theta_j)f(x,\theta)]$$
$$-2\,\Sigma_{\alpha=1}^{M}\,\Sigma_{\beta=1}^{M}[e_\alpha + f_\alpha(x,\theta^\circ) - f_\alpha(x,\theta)]\,\sigma^{\alpha\beta}(\tau)(\partial^2/\partial\theta_i\partial\theta_j)f_\beta(x,\theta)$$

$$\{(\partial/\partial\theta_i)s[Y(e,x,\theta^\circ),x,\tau,\theta]\}\{(\partial/\partial\theta_j)s[Y(e,x,\theta^\circ),x,\tau,\theta]\}$$
$$= 4[(\partial/\partial\theta_i)f(x,\theta)]'\Sigma^{-1}(\tau)[e + f(x,\theta^\circ) - f(x,\theta)]$$
$$\times [e + f(x,\theta^\circ) - f(x,\theta)]'\Sigma^{-1}(\tau)[(\partial/\partial\theta_j)f(x,\theta)]$$

to within a multiplicative constant. Since (Problem 4, Section 6),

$$(\partial/\partial\tau_i)\Sigma^{-1}(\tau) = -\Sigma^{-1}(\tau)[(\partial/\partial\tau_i)\Sigma(\tau)]\Sigma^{-1}(\tau)$$

we have

$$(\partial^2/\partial\tau_i\partial\theta_j)s[Y(e,x,\theta^\circ),x,\tau,\theta]$$

$$= 2[e + f(x,\theta^\circ) - f(x,\theta)]'\Sigma^{-1}(\tau)[(\partial/\partial\tau_i)\Sigma(\tau)]\Sigma^{-1}(\tau)(\partial/\partial\theta_j)f(x,\theta) .$$

Evaluating at $\theta = \theta^\circ = \theta^*$ and integrating we have

$$\int_{\mathcal{X}}\int_{\mathcal{E}}(\partial^2/\partial\tau_i\partial\theta_j)s[Y(e,x,\theta^*),x,\tau^*,\theta^*]\, dP(e)\, d\mu(x)$$

$$= \int_{\mathcal{X}} -2 \int_{\mathcal{E}} e\, dP(e)\Sigma^{-1}(\tau^*)[(\partial/\partial\tau_i)\Sigma(\tau^*)]\Sigma^{-1}(\tau^*)(\partial/\partial\theta_j)f(x,\theta^*)\, d\mu(x)$$

$$= 0$$

because $\int_{\mathcal{E}} e\, dP(e) = 0$. Thus, Assumption 6′ is enough to imply Assumption 6 of Chapter 3.

The parameters of the asymptotic distribution of $\hat{\theta}_n$ and various test statistics are defined in terms of the following.

<u>NOTATION 2</u>.

$$\Omega = \int_{\mathcal{X}}[(\partial/\partial\theta')f(x,\theta^*)]'\Sigma^{-1}[(\partial/\partial\theta')f(x,\theta^*)]\, d\mu(x)$$

$$\Omega_n^\circ = (1/n)\Sigma_{t=1}^n[(\partial/\partial\theta')f(x_t,\theta_n^\circ)]'\Sigma^{-1}[(\partial/\partial\theta')f(x_t,\theta_n^\circ)]$$

$$\Omega_n^* = (1/n)\Sigma_{t=1}^n[(\partial/\partial\theta')f(x_t,\theta_n^*)]'\Sigma^{-1}[(\partial/\partial\theta')f(x_t,\theta_n^*)]$$

NOTATION 3.

$\mathscr{I}^* = 4\Omega$

$\mathscr{J}^* = 2\Omega$

$\mathscr{U}^* = 0$

$\mathscr{I}_n^\circ = 4\Omega_n^\circ$

$\mathscr{J}_n^\circ = 2\Omega_n^\circ$

$\mathscr{U}_n^\circ = 0$

$\mathscr{I}_n^* = 4\Omega_n^*$

$\mathscr{J}_n^* = 2\Omega_n^* - (2/n)\Sigma_{t=1}^n \Sigma_{\alpha=1}^n \Sigma_{\beta=1}^n [f_\alpha(x_t,\theta_n^\circ) - f_\alpha(x_t,\theta_n^*)]\ \sigma^{\alpha\beta}(\partial^2/\partial\theta\partial\theta')f_\beta(x_t,\theta_n^*)$

$$\mathscr{U}_n^* = (4/n)\Sigma_{t=1}^n [(\partial/\partial\theta')f(x_t,\theta_n^*)]'\Sigma^{-1}[f(x_t,\theta_n^\circ) - f(x_t,\theta_n^*)]$$
$$\times [f(x_t,\theta_n^\circ) - f(x_t,\theta_n^*)]'\Sigma^{-1}[(\partial/\partial\theta')f(x_t,\theta_n^*)]$$

One can see from Notation 3 that it would be enough to have an estimator of Ω to be able to estimate $\mathscr{I}^*$ and $\mathscr{J}^*$. Accordingly we propose the following.

NOTATION 4.

$\hat{\Omega} = (1/n)\Sigma_{t=1}^n [(\partial/\partial\theta')f(x_t,\hat{\theta}_n)]'\hat{\Sigma}^{-1}[(\partial/\partial\theta')f(x_t,\hat{\theta}_n)]$

$\tilde{\Omega} = (1/n)\Sigma_{t=1}^n [(\partial/\partial\theta')f(x_t,\tilde{\theta}_n)]'\hat{\Sigma}^{-1}[(\partial/\partial\theta')f(x_t,\tilde{\theta}_n)]$

Since $(\mathscr{J}^*)^{-1}\mathscr{I}^*(\mathscr{J}^*)^{-1} = \Omega^{-1}$, we have from Theorem 5 of Chapter 3 that

$$\sqrt{n}\,(\hat{\theta}_n - \theta_n^\circ) \xrightarrow{\mathscr{L}} N(0,\Omega^{-1}) ,$$

$\hat{\Omega}$ converges almost surely to Ω .

Assumptions 7 and 13 of Chapter 3, restated in context, read as follows.

ASSUMPTION 7′. (Pitman drift) The sequence θ_n° is chosen such that $\lim_{n\to\infty}\sqrt{n}\,(\theta_n^\circ - \theta_n^*) = \Delta$. Moreover, $h(\theta^*) = 0$. $\square$

ASSUMPTION 13′. The function $h(\theta)$ is a once continuously differentiable mapping of Θ into R^q. Its Jacobian $H(\theta) = (\partial/\partial\theta')h(\theta)$ has full rank (=q) at $\theta = \theta^*$. □

From these last two assumptions we obtain a variety of ancillary facts, notably that $\tilde{\theta}_n$ converges almost surely to θ^*, that $\tilde{\Omega}$ converges almost surely to Ω, and that (Problem 6)

$$\mathcal{J}_n^* = 2\,\Omega_n^* + O(1/\sqrt{n}) .$$

The next task is to apply Theorems 11, 14, and 15 of Chapter 3 to obtain Wald, "likelihood ratio," and Lagrange multiplier test statistics as well as non-central chi-square approximations to their distributions. In the next section, these statistics will be modified so that tables of the non-central F-distribution can be used as approximations. However, we shall not try to justify these modifications by deriving characterization theorems as we did in Chapter 4 to justify the approach taken in Chapter 1. Instead, we shall merely note that the statistics proposed in the next section are asymptotically equivalent (differ by terms of order $o_p(1)$ or less) to the statistics derived below and let that serve as a justification.

Consider testing

$$H\colon h(\theta_n^o) = 0 \quad \text{against} \quad A\colon h(\theta_n^o) \neq 0$$

where, recall, $h(\theta)$ is a q-vector with Jacobian $H(\theta) = (\partial/\partial\theta')h(\theta)$, $H(\theta)$ being a q by p matrix. Writing $\hat{h} = h(\hat{\theta}_n)$ and $\hat{H} = H(\hat{\theta}_n)$ and applying Theorem 11 of Chapter 3 we have that the Wald test statistic is

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

and that the distribution of W' can be approximated by the non-central chi-square distribution with q degrees of freedom and non-centrality parameter

$$\alpha = n\,h'(\theta_n^o)[H(\theta_n^o)(\Omega_n^o)^{-1}H'(\theta_n^o)]^{-1}h(\theta_n^o)/2.$$

Multivariate nonlinear least squares is an instance where $\mathcal{J}_n^* \neq \mathcal{I}_n^*$ but $\mathcal{I}_n^* = (1/2)\mathcal{J}_n^* + O(1/\sqrt{n})$ (Problem 6) whence the likelihood ratio test statistic is

$$L' = n[s_n(\tilde{\theta}_n) - s_n(\hat{\theta}_n)] \ .$$

In the notation of the previous section,

$$L' = S(\tilde{\theta}_n, \hat{\Sigma}_n) - S(\hat{\theta}_n, \hat{\Sigma}_n) \ .$$

It is critical that $\hat{\Sigma}_n$ be the same in both terms on the right hand side of this equation. If they differ then the distributional results that follow are invalid (Problem 8). This seems a bit strange because it is usually the case in asymptotic theory that any $\sqrt{n}$ - consistent estimator of a nuisance parameter can be substituted in an expression without changing the result. The source of the difficulty is that the first equation in the proof of Theorem 15 is not true if $\hat{\Sigma}_n$ is not the same in both terms.

Applying, Theorem 15 of Chapter 3 and the remarks that follow it we have that the distribution of L' can be approximated by the non-central chi-square with q-degrees of freedom and non-centrality parameter (Problem 7)

$$\alpha = \{\Sigma_{t=1}^n [f(x_t, \theta_n^\circ) - f(x_t, \theta_n^*)]'\Sigma^{-1}(\partial/\partial\theta')f(x_t, \theta_n^*)\}$$
$$\times\, \Omega_n^{*-1}\{\Sigma_{t=1}^n [f(x_t, \theta_n^\circ) - f(x_t, \theta_n^*)]'\Sigma^{-1}(\partial/\partial\theta')f(x_t, \theta_n^*)\}'/(2n)$$

which is the same as for the efficient score test.

Up to this point, we have assumed that a correctly centered estimator of the variance-covariance matrix of the errors is available. That is, we have assumed that the estimator $\hat{\Sigma}_n$ has $\sqrt{n}\ (\hat{\Sigma}_n - \Sigma)$ bounded in probability whether $h(\theta_n^\circ) = 0$ is true or not. In the next section we shall see that this assumption is unrealistic with respect to the Lagrange multiplier test.

Accordingly, we base the Lagrange multiplier test on an estimate of scale $\hat{\Sigma}_n$ for which we assume that

$$\sqrt{n}(\tilde{\Sigma}_n - \Sigma_n^*) \text{ bounded in probability}$$

$$\ell im_{n\to\infty} \Sigma_n^* = \Sigma .$$

Of such estimators, that which is most likely to be used in applications is obtained by computing $\tilde{\theta}_n^{\#}$ to minimize $s_n(\theta,I)$ subject to $h(\theta) = 0$, where

$$s_n(\theta,V) = (1/n) \Sigma_{t=1}^n [y_t - f(x_t,\theta)]'V^{-1}[y_t-f(x_t,\theta)],$$

and then putting

$$\tilde{\tilde{\Sigma}}_n = (1/n) \Sigma_{t=1}^n[y_t - f(x_t,\tilde{\theta}_n^{\#})][y_t - f(x_t,\tilde{\theta}_n^{\#})]' .$$

The center is found (Problem 10) by computing $\theta_n^{\#}$ to minimize $s_n^\circ(\theta,I)$ subject to $h(\theta) = 0$ where

$$s_n^\circ(\theta,V) = \operatorname{tr} V^{-1}\Sigma + (1/n) \Sigma_{t=1}^n [f(x_t,\theta_n^\circ) - f(x_t,\theta)]'V^{-1}[f(x_t,\theta_n^\circ) - f(x_t,\theta)]$$

and putting

$$\Sigma_n^* = \Sigma + (1/n) \Sigma_{t=1}^n[f(x_t,\theta_n^\circ) - f(x_t,\theta_n^{\#})][f(x_t,\theta_n^\circ) - f(x_t,\theta_n^{\#})]' .$$

Using the estimator $\tilde{\Sigma}_n$, the formulas for the constrained estimators are revised to read $\tilde{\tilde{\theta}}_n$ minimizes $s_n(\theta,\tilde{\Sigma}_n)$ subject to $h(\theta) = 0$ and

$$\tilde{\tilde{\Omega}}_n = (1/n) \Sigma_{t=1}^n[(\partial/\partial\theta')f(x_t,\tilde{\tilde{\theta}}_n)]'\tilde{\Sigma}_n^{-1}[(\partial/\partial\theta')f(x_t,\tilde{\tilde{\theta}}_n)] .$$

The form of the efficient score or Lagrange multiplier test depends on how one goes about estimating V^* and $\mathcal{J}^*$ having the estimator $\tilde{\theta}_n$ in hand. In view of the remarks following Theorem 14 of Chapter 3, the choices

$$\tilde{\mathcal{V}} = \tilde{\Omega}^{-1}, \quad \tilde{\mathcal{J}} = 2\,\tilde{\Omega}$$

lead to considerable simplifications in the computations because it is not necessary to obtain second derivatives of $s_n(\theta)$ to estimate $\mathcal{J}^*$ and one is in the situation where $\tilde{\mathcal{J}}^{-1} = a\,\tilde{\mathcal{V}}$ for $a = 1/2$. With these choices, the efficient score test becomes (Problem 9)

$$R' = (1/n)\{\Sigma_{t=1}^{n}[y_t - f(x_t,\tilde{\tilde{\theta}}_n)]'\tilde{\Sigma}_n^{-1}(\partial/\partial\theta')f(x_t,\tilde{\tilde{\theta}}_n)\}$$
$$\times\ \tilde{\tilde{\Omega}}^{-1}\{\Sigma_{t=1}^{n}[y_t - f(x_t,\tilde{\tilde{\theta}}_n)]'\tilde{\Sigma}_n^{-1}(\partial/\partial\theta')f(x_t,\tilde{\tilde{\theta}}_n)\}' .$$

Let θ_n^{**} minimize $s^\circ(\theta,\Sigma_n^*)$ subject to $h(\theta) = 0$ and put

$$\Omega_n^{**} = (1/n)\ \Sigma_{t=1}^{n}[(\partial/\partial\theta')f(x_t,\theta_n^{**})]'(\Sigma_n^*)^{-1}[(\partial/\partial\theta')f(x_t,\theta_n^{**})] .$$

Then the distribution of $\tilde{R}'$ can be characterized as (Problem 9)

$$\tilde{R}' = \tilde{Y} + o_p(1)$$

where

$$\tilde{Y} = \tilde{Z}'\left(\Omega_n^{**}\right)^{-1}H_n^{**\prime}[H_n^{**}(\Omega_n^{**})^{-1}H_n^{**\prime}]^{-1}H_n^{**}\left(\Omega_n^{**}\right)^{-1}\tilde{Z}$$

and

$$\tilde{Z} \sim N\{(-1/\sqrt{n})\Sigma_{t=1}^{n}[(\partial/\partial\theta')f(x_t,\theta_n^{**})]'(\Sigma_n^*)^{-1}[f(x_t,\theta_n^\circ) - f(x_t,\theta_n^{**})], \mathcal{J}_n^{**}/4\}$$

$$\mathcal{J}_n^{**}/4 = (1/n)\Sigma_{t=1}^{n}[(\partial/\partial\theta')f(x_t,\theta_n^{**})]'(\Sigma_n^*)^{-1}\Sigma(\Sigma_n^*)^{-1}[(\partial/\partial\theta')f(x_t,\theta_n^{**})]$$

$$H_n^{**} = (\partial/\partial\theta')\ h(\theta_n^{**}) .$$

The random variable $\tilde{Y}$ is a general quadratic form in the multivariate normal random variable $\tilde{Z}$ and one can use the methods discussed in Imhof (1961) to compute its distribution. Comparing with the result derived in Problem 5 one sees that the unfortunate consequence of the use of $\tilde{\Sigma}_n$ instead of $\hat{\Sigma}_n$ to compute $\tilde{R}'$ is that one can not use tables of the non-central chi-square to approximate its non-null distribution. The null distribution of $\tilde{Y}$ is a chi-square with q degrees of freedom since, if the null is true, $\theta_n^{**} = \theta_n^o$ and and $\Sigma_n^* = \Sigma$.

PROBLEMS

1. Show that the regularity conditions listed in this section are sufficient to imply the regularity conditions listed in Section 2 of Chapter 4 for each of the models

$$y_\alpha = f_\alpha(\theta_\alpha) + e_\alpha \qquad \alpha = 1,2,\ldots, M .$$

Show that this implies that $\hat{\theta}^{\#}_\alpha$ converges almost surely to θ^*_α and that $\sqrt{n}\,(\hat{\theta}^{\#}_\alpha - \theta^\circ_{\alpha n})$ is bounded in probability where θ^*_α and $\theta^\circ_{\alpha n}$ are defined by Assumption 4′ using

$$\theta' = (\theta_1', \; \theta_2', \ldots, \theta_\alpha', \ldots, \theta_M') .$$

Let

$$\sigma_{\alpha\beta}(\theta) = (1/n)[y_\alpha - f_\alpha(\theta_\alpha)]'[y_\beta - f_\beta(\theta_\beta)] .$$

Show that $\sigma_{\alpha\beta}(\hat{\theta}^{\#})$ converges almost surely to $\sigma_{\alpha\beta}(\theta^*)$ and that $\sqrt{n}\,[\sigma_{\alpha\beta}(\hat{\theta}^{\#}) - \sigma_{\alpha\beta}(\theta^\circ_n)]$ is bounded in probability.

Hint. Use Taylor's theorem to write

$$\begin{aligned}\sqrt{n}\,[\sigma_{\alpha\beta}(\hat{\theta}^{\#}) - \sigma_{\alpha\beta}(\theta^\circ_n)] &= \\ &= \{(-1/n)\Sigma_{t=1}^n[y_{\alpha t} - f(x_t,\bar{\theta})](\partial/\partial\theta')f_\beta(x_t,\bar{\theta})\}\,\sqrt{n}\,(\hat{\theta} - \theta^\circ_n) \\ &\quad + \{(-1/n)\Sigma_{t=1}^n[y_{\beta t} - f(x_t,\bar{\theta})](\partial/\partial\theta')f_\alpha(x_t,\bar{\theta})\}\,\sqrt{n}\,(\hat{\theta} - \theta^\circ_n) .\end{aligned}$$

2. Apply Lemma 2 of Chapter 3 to conclude that one can assume that $\hat{\tau}_n$ is in T without loss of generality.

3. Use $\Sigma^{-1}(\tau)$ = adjoint $\Sigma(\tau)/\det\Sigma(\tau)$ to show that $\Sigma^{-1}(\tau)$ is continuous and differentiable over T .

4. Verify the computations of Notation 3.

5. (Lagrange multiplier test with a correctly centered estimator of scale). Assume that an extimator $\hat{\Sigma}$ is available with $\sqrt{n}\,(\hat{\Sigma}_n - \Sigma)$ bounded in probability. Apply Theorem 14 with modifications as necessary to reflect the choices $\tilde{V} = \tilde{\Omega}^{-1}$ and $\tilde{\mathcal{J}} = 2\,\tilde{\Omega}$ to conclude that the efficient score test statistic is

$$R' = (1/n)\{\Sigma_{t=1}^{n}[y_t - f(x_t,\tilde{\theta}_n)]'\hat{\Sigma}^{-1}(\partial/\partial\theta')f(x_t,\tilde{\theta}_n)\}\,\tilde{\Omega}^{-1}\tilde{H}'(\tilde{H}\,\tilde{\Omega}^{-1}\tilde{H}')^{-1}\tilde{H}\,\tilde{\Omega}^{-1}$$

$$\times\,\{\Sigma_{t=1}^{n}[y_t - f(x_t,\tilde{\theta}_n)]'\hat{\Sigma}^{-1}(\partial/\partial\theta')f(x_t,\tilde{\theta}_n)\}'$$

where $\tilde{H} = H(\tilde{\theta}_n)$ with a distribution that can be characterized as

$$R' = Y + o_p(1)$$

where

$$Y = Z'\Omega_n^{*-1}H_n^{*\prime}[H_n^{*}\,\Omega_n^{*-1}H_n^{*\prime}]^{-1}\,H_n^{*}\,\Omega_n^{*-1}Z$$

and

$$Z \sim N\{(-1/\sqrt{n})\Sigma_{t=1}^{n}[(\partial/\partial\theta')f(x_t,\theta_n^{*})]'\Sigma^{-1}[f(x_t,\theta_n^{\circ}) - f(x_t,\theta_n^{*})],\ \Omega_n^{*}\}$$

with $H_n^{*} = H(\theta_n^{*})$. Show that the random variable Y has the noncentral chi-square distribution with q degrees of freedom and noncentrality parameter

$$\alpha = \{\Sigma_{t=1}^{n}[f(x_t,\theta_n^{\circ}) - f(x_t,\theta_n^{*})]'\Sigma^{-1}(\partial/\partial\theta')f(x_t,\theta_n^{*})\}$$

$$\times\,\Omega_n^{*-1}\,H_n^{*\prime}[H_n^{*}\,\Omega_n^{*-1}H_n^{*\prime}]H_n^{*}\,\Omega_n^{*-1}$$

$$\times\,\{\Sigma_{t=1}^{n}[f(x_t,\theta_n^{\circ}) - f(x_t,\theta_n^{*})]'\Sigma^{-1}(\partial/\partial\theta')f(x_t,\theta_n^{*})\}'/(2n)\ .$$

Now use the fact that $\tilde{V} = 1/2\,\tilde{\mathcal{J}}^{-1}$ to obtain the simpler form

$$R' = (1/n)\{\Sigma_{t=1}^{n}[y_t - f(x_t,\tilde{\theta}_n)]'\hat{\Sigma}^{-1}(\partial/\partial\theta')f(x_t,\tilde{\theta}_n)\}$$

$$\times\,\tilde{\Omega}^{-1}\{\Sigma_{t=1}^{n}[y_t - f(x_t,\tilde{\theta}_n)]'\hat{\Sigma}^{-1}(\partial/\partial\theta')f(x_t,\tilde{\theta}_n)\}'\ .$$

Use the same sort of argument to show that

$$\alpha = \{\Sigma_{t=1}^{n}[f(x_t,\theta_n^o) - f(x_t,\theta_n^*)]'\Sigma^{-1}(\partial/\partial\theta')f(x_t,\theta_n^*)\}$$
$$\times\, \Omega_n^{*-1}\{\Sigma_{t=1}^{n}[f(x_t,\theta_n^o) - f(x_t,\theta_n^*)]'\Sigma^{-1}(\partial/\partial\theta')f(x_t,\theta_n^*)\}'/(2n) \ .$$

Hint. See the remarks and the example which follow the proof of Theorem 14 of Chapter 3.

6. Verify that $\mathcal{J}_n^* = (1/2)\mathcal{I}_n^* + o(1/\sqrt{n})$. Hint. See the example following Theorem 15 of Chapter 3.

7. Show that

$$[(\partial/\partial\theta)s_n^o(\theta_n^*)]'\Omega_n^{*-1}H_n^{*\prime}(H_n^*\Omega_n^{*-1}H_n^*)^{-1}H_n^*\Omega_n^{*-1}[(\partial/\partial\theta)s_n^o(\theta_n^*)]$$
$$= [(\partial/\partial\theta)s_n^o(\theta_n^*)]'\Omega_n^{*-1}[(\partial/\partial\theta)s_n^o(\theta_n^*)] \ .$$

Hint. See Problem 5.

8. Suppose that $\hat{\hat{\Sigma}}_n$ is computed as

$$\hat{\hat{\Sigma}}_n = (1/n)\Sigma_{t=1}^{n}[y_t - f(x_t,\hat{\hat{\theta}}_n)][y_t - f(x_t,\hat{\theta}_n)]'$$

and that $\tilde{\Sigma}_n$ is computed as

$$\tilde{\Sigma}_n = (1/n)\Sigma_{t=1}^{n}[y_t - f(x_t,\tilde{\theta}_n)][y_t - f(x_t,\tilde{\theta}_n)]' \ .$$

Take it as given that both $\hat{\Sigma}_n$ and $\tilde{\Sigma}_n$ converge almost surely to Σ and that both $\sqrt{n}\,(\hat{\Sigma}_n - \Sigma)$ and $\sqrt{n}(\tilde{\Sigma}_n - \Sigma)$ are bounded in probability. Show that both $S(\hat{\theta}_n,\hat{\hat{\Sigma}}_n) \equiv M$ and $S(\tilde{\theta}_n,\tilde{\Sigma}_n) \equiv M$ so that

$$S(\tilde{\theta}_n,\tilde{\Sigma}_n) - S(\hat{\theta}_n,\hat{\hat{\Sigma}}_n) \equiv 0$$

and cannot be asymptotically distributed as a chi-square random variable. However, both

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

and

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

are asymptotically distributed as a chi-square random variable by the results of this section.

9. (Lagrange multiplier test with a miscentered estimator of scale). Suppose that one uses an estimator of scale $\tilde{\Sigma}_n$ with $\sqrt{n}\,(\tilde{\Sigma}_n - \Sigma_n^*)$ bounded in probability and $\ell im_{n\to\infty}\,\Sigma_n^* = \Sigma$ as in the text. Use the same argument as in Problem 5 to show that the choices $\tilde{V} = \tilde{\tilde{\Omega}}^{-1}$ and $\tilde{\mathcal{J}} = 2\,\tilde{\tilde{\Omega}}$ allow the Lagrange multiplier test to be written as

$$\tilde{R}' = (1/n)\{\Sigma_{t=1}^n[y_t - f(x_t,\tilde{\theta}_n)]'\tilde{\Sigma}_n^{-1}(\partial/\partial\theta')f(x_t,\tilde{\theta}_n)\}$$

$$X\ \tilde{\Omega}_n^{-1}\{\Sigma_{t=1}^n[y_t - f(x_t,\tilde{\theta}_n)]'\tilde{\Sigma}_n^{-1}(\partial/\partial\theta')f(x_t,\tilde{\theta}_n)\}' .$$

Show that the distribution of $\tilde{R}'$ can be characterized as $\tilde{R}' = \tilde{Y} + o_p(1)$ with $\tilde{Y}$ as given in the text.

Hint. Let $H = H_n^{**}$, $\mathcal{I} = \mathcal{I}_n^{**}$, $\mathcal{J} = (\partial^2/\partial\theta\partial\theta')s_n^\circ(\theta_n^{**}, \Sigma_n^*)$, and $\Omega = \Omega_n^{**}$. Note that $\mathcal{J} = 2\,\Omega_n^{**} + o(1)$. Use Theorems 12 and 13 of Chapter 3 to show that

$$
\begin{aligned}
R' &= (1/4)[\sqrt{n}(\partial/\partial\theta)s_n(\tilde{\tilde{\theta}}_n)]'\tilde{\tilde{\Omega}}_n^{-1}[\sqrt{n}(\partial/\partial\theta)s_n(\tilde{\tilde{\theta}}_n)] \\
&= (1/4)[\sqrt{n}(\partial/\partial\theta)s_n(\tilde{\tilde{\theta}}_n)]'\Omega^{-1}[\sqrt{n}(\partial/\partial\theta)s_n(\tilde{\tilde{\theta}}_n)] + o_p(1) \\
&= (1/4)[\sqrt{n}(\partial/\partial\theta)s_n(\theta_n^{**})]'\mathcal{J}^{-1}H'(H\mathcal{J}^{-1}H')^{-1}H\,\Omega^{-1}H'(H\mathcal{J}^{-1}H')^{-1} \\
&\qquad H\mathcal{J}^{-1}[\sqrt{n}(\partial/\partial\theta)s_n(\theta^{**})] + o_p(1) \\
&= (1/4)[\sqrt{n}(\partial/\partial\theta)s_n(\theta_n^{**})]'\Omega^{-1}H'(H\,\Omega^{-1}H')^{-1}H\Omega^{-1}[\sqrt{n}(\partial/\partial\theta)s_n(\theta^{**})] + o_p(1) \\
&= (X/2)'\Omega^{-1}H'(H\,\Omega^{-1}H')^{-1}H\,\Omega^{-1}(X/2) + o_p(1)
\end{aligned}
$$

where

$$X \sim [\sqrt{n}(\partial/\partial\theta)s_n^\circ(\theta_n^{**})/2,\ \mathcal{I}_n^{**}/4] .$$

10. (Computation of the value Σ_n^* that centers $\tilde{\Sigma}_n$). Assume that $h(\theta) = 0$ can be written equivalently as $\theta = g(\rho)$ for some ρ. Use Theorem 5 of Chapter 3 to show that if one computes $\hat{\rho}_n$ to minimize

$$s_n(\rho) = (1/n)\Sigma_{t=1}^n \{y_t - f[x_t, g(\rho)]\}'\{y_t - f[x_t, g(\rho)]\}$$

then the appropriate centering value is computed by finding ρ_n° to minimize

$$s_n^\circ(\rho) = (1/n)\int_{\mathcal{E}} \{e + f(x_t, \theta_n^\circ) - f[x_t, g(\rho)]\}'\{e + f(x_t, \theta_n^\circ) - f[x_t, g(\rho)]\} dP(e$$

$$= \operatorname{tr} \Sigma + (1/n)\Sigma_{t=1}^n \{f(x_t, \theta_n^\circ) - f[x_t, g(\rho)]\}'\{f(x_t, \theta_n^\circ) - f[x_t, g(\rho)]\} .$$

Now

$$\tilde{\Sigma}_n = (1/n)\Sigma_{t=1}^n \{y_t - f[x_t, g(\hat{\rho}_n)]\}\{y_t - f[x_t, g(\hat{\rho}_n)]\}'$$

is the solution of the following minimization problem (Problem 11)

minimize: $s_n(V, \hat{\rho}_n) = (1/n)\Sigma_{t=1}^n \ell n \det(V) + \{y_t - f[x_t, g(\hat{\rho}_n)]\}' V^{-1}\{y_t - f[x_t, g(\hat{\rho}_n)]$

subject to: V positive definite, symmetric.

Use Theorem 5 of Chapter 3 to show that the value Σ_n^* that centers $\tilde{\Sigma}_n$ is computed as the solution of the problem

minimize: $s_n^\circ(V, \rho_n^\circ)$

subject to: V positive definite, symmetric

where

$$s_n^\circ(V,\rho) = \ell n \det(V) + (1/n)\Sigma_{t=1}^n \int_{\mathcal{E}} \{e + f(x_t,\theta_n^\circ) - f[x_t,g(\rho)]\}' V^{-1} \{e + f(x_t,\theta_n^\circ) - f[x_t,g(\rho)]\} dP(e)$$

$$= \ell n \det(V) + \mathrm{tr}(V^{-1}\Sigma) + (1/n)\Sigma_{t=1}^n \{f(x_t,\theta_n^\circ) - f[x_t,g(\rho)]\}' V^{-1} \{f(x_t,\theta_n^\circ) - f[x_t,g(\rho)]\} .$$

The solution of this minimization problem is (Problem 11)

$$\Sigma_n^\circ = \Sigma + (1/n)\Sigma_{t=1}^n \{f(x_t,\theta_n^\circ) - f[x_t,g(\bar{\rho}_n^\circ)]\}\{f(x_t,\theta_n^\circ) - f[x_t,g(\bar{\rho}_n^\circ)]\}' .$$

11. Let

$$f(V) = \ell n \det V + \mathrm{tr}\,(V^{-1}A)$$

where A is an M by M positive definite symmetric matrix. Show that the minimum of f(V) over the (open) set of all positive definite matrices V is attained at the value V = A. Hint.

$$f(V) - f(A) = -\ell n \det(V^{-1}A) + \mathrm{tr}\,(V^{-1}A) - M .$$

Let λ_i be the eigenvalues of $V^{-1}A$. Then

$$f(V) - f(A) = \Sigma_{i=1}^M (\lambda_i - \ell n \lambda_i - 1) .$$

Since the line $y = x$ plots above the line $y = \ell n\, x + 1$ one has

$$f(V) - f(A) > 0 \quad \text{if any } \lambda_i \neq 1$$

$$f(V) - f(A) = 0 \quad \text{if all } \lambda_i = 1 .$$

12. (Efficiency of least squares estimators) Define $\hat{\theta}^{\#}$ as the minimizer of $s_n(\theta,\hat{V}_n)$ where $\sqrt{n}(\hat{V}_n - V)$ is bounded in probability, $\ell im_{n\to\infty} \hat{V}_n = V$ almost surely, and V is positive definite. Show that under Assumptions 1′ through 7′

$$\sqrt{n}(\hat{\theta}^{\#} - \theta_n^{\circ}) \xrightarrow{\mathcal{L}} N(0, \mathcal{J}_V^{-1} \mathcal{I}_V \mathcal{J}_V^{-1})$$

with

$$\mathcal{J}_V = 2\int_{\mathcal{X}} [(\partial/\partial\theta')f(x,\theta^*)]' \, V^{-1}[(\partial/\partial\theta')f(x,\theta^*)]d\mu(x)$$

$$\mathcal{I}_V = 4\int_{\mathcal{X}} [(\partial/\partial\theta')f(x,\theta^*)]' \, V^{-1}\Sigma \, V^{-1}[(\partial/\partial\theta')f(x,\theta^*)]d\mu(x) \; .$$

Show that $a'\mathcal{J}_V^{-1}\mathcal{I}_V\mathcal{J}_V^{-1}a$ is minimized when $V = \Sigma$. Note that the equation-by-equation estimator has $V = I$.

4. HYPOTHESIS TESTING

The results of this section are expressed in a summation notation using the multivariate notational scheme. A summation notation is less attractive aesthetically than a matrix notation using Kroneker products but formulas written in summation notation translate easily into machine code, as noted earlier, and have pedagogical advantages.
At the end of the section is a summary of the results using a matrix notation for data arranged in the "seemingly unrelated" scheme.
Assume that the data follow the model

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

with the functional form $f(x,\theta)$ known, x_t a k-vector, θ a p-vector, y_t an M-vector, and e_t an M-vector. Assume that the errors $\{e_t\}$ are independently and normally distributed each with mean zero and variance-covariance matrix Σ. The unknown parameters are θ° and Σ.

Consider testing a hypothesis that can be expressed either as a parametric restriction

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

or as a functional dependency

$$H\colon\ \theta^\circ = g(\rho^\circ) \text{ for some } \rho^\circ \text{ against } A\colon\ \theta^\circ \neq g(\rho) \text{ for any } \rho.$$

Here, $h(\theta)$ maps $\mathbb{R}^p$ into $\mathbb{R}^q$ with Jacobian

$$H(\theta) = (\partial/\partial\theta')\, h(\theta)$$

which we assume is continuous and has rank q at θ° the true value of θ; $g(\rho)$

maps $\mathbb{R}^r$ into $\mathbb{R}^p$ and has Jacobian

$$G(\rho) = (\partial/\partial\rho')g(\rho) .$$

The Jacobians are of order q by p for $H(\theta)$ and p by r for $G(\rho)$; we assume that $p = r + q$ and from $h[g(\rho)] = 0$ we have $H[g(\rho)]G(\rho) = 0$. For complete details see Section 6 of Chapter 3. Let us illustrate with the example.

EXAMPLE 1 (Continued) Recall that the model

$$y_t = f(x_t, \theta^\circ) + e_t \qquad t = 1,2,\ldots,224$$

with

$$f(x,\theta) = \begin{pmatrix} \ln \dfrac{a_1 + b_{11}x_1 + b_{12}x_2 + b_{13}x_3}{-1 + b_{13}x_1 + b_{23}x_2 + b_{33}x_3} \\[2ex] \ln \dfrac{a_2 + b_{12}x_1 + b_{22}x_2 + b_{23}x_3}{-1 + b_{13}x_1 + b_{23}x_2 + b_{33}x_3} \end{pmatrix}$$

$$\theta' = (a_1, b_{11}, b_{12}, b_{13}, a_2, b_{22}, b_{23}, b_{33})$$

was chosen as a reasonable representation of the data of Table 1 on the basis of the computations reported in Table 3. Since we have settled on a model specification, let us henceforth adopt the simpler subscripting scheme

$$f(x,\theta) = \begin{pmatrix} \ln \dfrac{\theta_1 + \theta_2x_1 + \theta_3x_2 + \theta_4x_3}{-1 + \theta_4x_1 + \theta_7x_2 + \theta_8x_3} \\[2ex] \ln \dfrac{\theta_5 + \theta_3x_1 + \theta_6x_2 + \theta_7x_3}{-1 + \theta_4x_1 + \theta_7x_2 + \theta_8x_3} \end{pmatrix}$$

$$\theta = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8)' .$$

In this notation, the hypothesis of homogeneity may be written as the parametric restriction

$$h(\theta) = \begin{pmatrix} \theta_2 + \theta_3 + \theta_4 \\ \theta_3 + \theta_6 + \theta_7 \\ \theta_4 + \theta_7 + \theta_8 \end{pmatrix} = 0$$

with Jacobian

$$H(\theta) = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \end{pmatrix} .$$

The hypothesis may also be written as a functional dependency

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \\ \theta_6 \\ \theta_7 \\ \theta_8 \end{pmatrix} = \begin{pmatrix} \theta_1 \\ -\theta_3 - \theta_4 \\ \theta_3 \\ \theta_4 \\ \theta_5 \\ -\theta_7 - \theta_3 \\ \theta_7 \\ -\theta_4 - \theta_7 \end{pmatrix} = \begin{pmatrix} \rho_1 \\ -\rho_2 - \rho_3 \\ \rho_2 \\ \rho_3 \\ \rho_4 \\ -\rho_5 - \rho_2 \\ \rho_5 \\ -\rho_5 - \rho_3 \end{pmatrix} = g(\rho)$$

with Jacobian

$$G(\rho) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 & -1 \end{pmatrix}$$

which is, of course, the same as was obtained in Section 2. In passing, observe that $H[g(\rho)]G(\rho) = 0$. []

Throughout this section we shall take $\hat{\Sigma}$ to be any random variable that converges almost surely to Σ and has $\sqrt{n}(\hat{\Sigma} - \Sigma)$ bounded in probability. To obtain a level α test this condition on $\hat{\Sigma}$ need only hold when H: $h(\theta^\circ) = 0$ is true but to use the approximations to power derived below the condition must hold when A: $h(\theta^\circ) \neq 0$ as well.

There are two commonly used estimators of Σ that satisfy the condition under both the null and alternative hypotheses. We illustrated one of them in Section 2. There one fitted each equation in the model separately in the style of Chapter 1 and then estimated Σ from single equation residuals. Recalling that

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

an alternative approach is to put $\Sigma = I$, minimize $S(\theta,I)$ with respect to θ to obtain $\hat{\theta}^{\#}$, and estimate Σ by

$$\hat{\Sigma} = (1/n)\ \Sigma_{t=1}^{n}[y_t - f(x_t, \hat{\theta}^{\#})][y_t - f(x_t,\hat{\theta}^{\#})]'$$

$$= (1/n)\ \Sigma_{t=1}^{n}\ \hat{e}_t\ \hat{e}_t' \quad .$$

If there are no across equation restrictions on the model these two estimators will be the same. When there are across equation restrictions, there is a tendency to incorporate them directly into the model specification when using the multivariate notational scheme as we have just done with the example. (The restrictions that θ_3, θ_4, θ_7, and θ_8 be the same in both equations are the across equation restrictions, a total of four. The restriction that θ_4 be the same in the numerator and denominator of the first equation is called a within equation restriction.) This tendency to incorporate across equation restrictions in the model specification causes the two estimators of Σ to be different in most instances. Simply for varieties sake, we shall use the estimator computed from the fit that minimizes $S(\theta,I)$ in this section.

We illustrate these ideas with the example. In reading what follows, recall the ideas used to write a multivariate model in a univariate notation. Factor $\hat{\Sigma}^{-1}$ as $\hat{\Sigma}^{-1} = \hat{P}'\hat{P}$, let $\hat{p}'_{(\alpha)}$ denote a typical row of $\hat{P}$, and put

$$s = M(t-1) + \alpha$$

$$"y_s" = \hat{p}'_{(\alpha)}\, y_t$$

$$"x_s" = (\hat{p}'_{(\alpha)},\, x'_t)'$$

$$"f"("x_s",\theta) = \hat{p}'_{(\alpha)}\, f(x_t,\theta)\ .$$

In this notation

$$S(\theta,\hat{\Sigma}) = \Sigma_{s=1}^{nM}\, ["y_s" - "f"("x_s",\theta)]^2\ .$$

EXAMPLE 1. (continued) SAS code to minimize $S(\theta,I)$ for

$$f(x,\theta) = \begin{pmatrix} \ell n\ \dfrac{\theta_1 + \theta_2 x_1 + \theta_3 x_2 + \theta_4 x_3}{-1 + \theta_4 x_1 + \theta_7 x_2 + \theta_8 x_3} \\ \\ \ell n\ \dfrac{\theta_5 + \theta_3 x_1 + \theta_6 x_2 + \theta_7 x_3}{-1 + \theta_4 x_1 + \theta_7 x_2 + \theta_8 x_3} \end{pmatrix}\ .$$

is shown in Figure 3a. A detailed discussion of the ideas is found in connection with Figure 2a; briefly they are as follows.

Trivially the identity factors as $I = P'P$ with $P = I$. The multivariate observations y_t, x_t for $t = 1, 2, \ldots, 224 = n$ are transformed to the univariate entities

Figure 3a. Example 1 Fitted by Least Squares, Across Equation Constraints Imposed

SAS Statements:

```
DATA WORK01;  SET EXAMPLE1;
P1=1.0;  P2=0.0;  Y=P1*Y1+P2*Y2;  OUTPUT;
P1=0.0;  P2=1.0;  Y=P1*Y1+P2*Y2;  OUTPUT;  DELETE;
PROC NLIN DATA=WORK01 METHOD=GAUSS ITER=50 CONVERGENCE=1.E-13;
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;
MODEL Y=P1*LOG(PEAK/BASE)+P2*LOG(INTER/BASE);
DER.T1=P1/PEAK;  DER.T2=P1/PEAK*X1;  DER.T3=P1/PEAK*X2+P2/INTER*X1;
DER.T4=P1/PEAK*X3+(-P1-P2)/BASE*X1;  DER.T5=P2/INTER;
DER.T6=P2/INTER*X2;  DER.T7=P2/INTER*X3+(-P1-P2)/BASE*X2;
DER.T8=(-P1-P2)/BASE*X3;
OUTPUT OUT=WORK02 RESIDUAL=E;
```

Output:

```
                                      SAS                                        1

                    NON-LINEAR LEAST SQUARES ITERATIVE PHASE

                DEPENDENT VARIABLE: Y          METHOD: GAUSS-NEWTON

   ITERATION            T1              T2              T3            RESIDUAL SS
                        T4              T5              T6
                        T7              T8

       0           -2.90000000     -1.30000000      0.82000000        68.32779625
                    0.36000000     -1.50000000     -1.00000000
                   -0.03000000     -0.47000000

       .
       .
       .

      14           -2.98025942     -1.16088895      0.78692676        57.02306899
                    0.35309087     -1.50604388     -0.99985707
                    0.05407441     -0.47436347

NOTE: CONVERGENCE CRITERION MET.
```

$$"y_s" = p'_{(\alpha)} y_t \, , \qquad "x_s" = (p'_{(\alpha)}, x'_t)'$$

for s = 1, 2, ..., 448 = nM which are then stored in the data set WORKØ1 as shown in Figure 3a. The univariate nonlinear model

$$"y_s" = "f"("x_s", \theta) + "e_s" \qquad s = 1, 2, \ldots, 448 = nM$$

with

$$"f"("x_s", \theta) = p'_{(\alpha)} f(x_t, \theta)$$

$$s = M(t-1) + \alpha$$

is fitted to these data using PROC NLIN and the residuals $"\hat{e}_s"$ for s = 1, 2, ..., 448 = nM are stored in the data set named WORKØ2.

In Figure 3b the univariate residuals stored in WORKØ2 are regrouped into the multivariate residuals $\hat{e}_t$ for t = 1, 2, ..., 224 = n and stored in a data set named WORKØ5; here we are exploiting the fact that P = I. From the residuals stored in WORKØ5, $\hat{\Sigma}$ and $\hat{P}$ with $\hat{\Sigma}^{-1} = \hat{P}'\hat{P}$ are computed using PROC MATRIX. Compare this estimate of Σ with the one obtained in Figure 1c. Imposing the across equation restrictions results in a slight difference between the two estimates.

Using $\hat{P}$ as computed in Figure 3b, $S(\theta, \hat{\Sigma})$ is minimized to obtain

$$\hat{\theta} = \begin{pmatrix} -2.92458126 \\ -1.28674630 \\ 0.81856986 \\ 0.36115784 \\ -1.53758854 \\ -1.04895916 \\ 0.03008670 \\ -0.46742014 \end{pmatrix} \qquad \text{(from Figure 3c)}$$

as shown in Figure 3c; the ideas are the same as for Figure 3a. The difference between $\hat{\Sigma}$ in Figures 1c and 3b results in a slight difference between the estimate of θ computed in Figure 2c and $\hat{\theta}$ above. []

Figure 35. Contemporaneous Variance-Covariance Matrix of Example 1 Estimated from Least Squares Residuals, Across Equation Constraints Imposed.

SAS Statements:

```
DATA WORK03;  SET WORK02;   E1=E;  IF MOD(_N_,2)=0 THEN DELETE;
DATA WORK04;  SET WORK02;   E2=E;  IF MOD(_N_,2)=1 THEN DELETE;
DATA WORK05;  MERGE WORK03 WORK04;  KEEP E1 E2;
PROC MATRIX FW=20; FETCH E DATA=WORK05(KEEP=E1 E2);
SIGMA=E'*E#/224;  PRINT SIGMA;  P=HALF(INV(SIGMA));  PRINT P;
```

Output:

```
                                        SAS                                    4


SIGMA                 COL1                COL2
ROW1       0.1649246288351    0.09200572942276
ROW2      0.09200572942276    0.08964264342294


P                     COL1                COL2

ROW1         3.76639099219        -3.865677509
ROW2                     0       3.33997082052 4
```

Figure 3c. Example 1 Fitted by Multivariate Least Squares, Across Equation Constraints Imposed.

SAS Statements:

```
DATA WORK01;  SET EXAMPLE1;
P1=3.76639099219;  P2=-3.865677509;  Y=P1*Y1+P2*Y2;  OUTPUT;
P1=0.0;  P2=3.339970820524;  Y=P1*Y1+P2*Y2;  OUTPUT;  DELETE;
PROC NLIN DATA=WORK01 METHOD=GAUSS ITER=50 CONVERGENCE=1.E-13;
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;
MODEL Y=P1*LOG(PEAK/BASE)+P2*LOG(INTER/BASE);
DER.T1=P1/PEAK;  DER.T2=P1/PEAK*X1;  DER.T3=P1/PEAK*X2+P2/INTER*X1;
DER.T4=P1/PEAK*X3+(-P1-P2)/BASE*X1;  DER.T5=P2/INTER;
DER.T6=P2/INTER*X2;  DER.T7=P2/INTER*X3+(-P1-P2)/BASE*X2;
DER.T8=(-P1-P2)/BASE*X3;
```

Output:

```
                                    SAS                                      5

                  NON-LINEAR LEAST SQUARES ITERATIVE PHASE

             DEPENDENT VARIABLE: Y          METHOD: GAUSS-NEWTON

  ITERATION           T1            T2            T3          RESIDUAL SS
                      T4            T5            T6
                      T7            T8

      0          -2.90000000   -1.30000000    0.82000000    543.55788176
                  0.36000000   -1.50000000   -1.00000000
                 -0.03000000   -0.47000000
      .
      .
      .
     14          -2.92458126   -1.28674630    0.81856986    446.85695247
                  0.36115784   -1.53758854   -1.04895916
                  0.03008670   -0.46742014

NOTE: CONVERGENCE CRITERION MET.

                                    SAS                                      7

     NON-LINEAR LEAST SQUARES SUMMARY STATISTICS     DEPENDENT VARIABLE Y

       SOURCE                  DF       SUM OF SQUARES        MEAN SQUARE

       REGRESSION               8       6468.84819992        808.60602499
       RESIDUAL               440        446.85695247          1.01558398
       UNCORRECTED TOTAL      448       6915.70515239

       (CORRECTED TOTAL)      447        866.32697265

PARAMETER        ESTIMATE        ASYMPTOTIC               ASYMPTOTIC 95 %
                                 STD. ERROR             CONFIDENCE INTERVAL
                                                     LOWER            UPPER
T1            -2.92458126      0.27790948      -3.47078451      -2.37837801
T2            -1.28674630      0.22671234      -1.73232670      -0.84116589
T3             0.81856986      0.08088226       0.65960389       0.97753584
T4             0.36115784      0.03029057       0.30162474       0.42069093
T5            -1.53758854      0.09192958      -1.71826692      -1.35691016
T6            -1.04895916      0.08367724      -1.21341839      -0.88449993
T7             0.03008670      0.03614145      -0.04094570       0.10111909
T8            -0.46742014      0.01926170      -0.50527708      -0.42956320
```

The theory in Section 3 would lead one to test H: $h(\theta) = 0$ by computing, for instance,

$$L' = S(\tilde{\theta},\hat{\Sigma}) - S(\hat{\theta},\hat{\Sigma})$$

and rejecting if L' exceeds the α-level critical point of the χ^2-distribution with q degrees of freedom, recall that $\tilde{\theta}$ minimizes $S(\theta,\hat{\Sigma})$ subject to $h(\theta) = 0$ and that $\hat{\theta}$ is the unconstrained minimum of $S(\theta,\hat{\Sigma})$. This is what one usually encounters in the applied literature. We shall not use that approach here. In this instance we shall compute

$$L = \frac{[S(\tilde{\theta},\hat{\Sigma}) - S(\hat{\theta},\hat{\Sigma})]/q}{S(\hat{\theta},\hat{\Sigma})/(nM - p)}$$

and reject if L exceeds the α-level critical point of the F-distribution with q numerator degrees of freedom and $nM - p$ denominator degrees of freedom. There are two reasons for doing so. One is pedogogical, we wish to transfer the ideas in Chapter 1 intact to the multivariate setting. The other is to make some attempt to compensate for the sampling variation due to having to estimate Σ. We might note that $S(\hat{\theta}^{\#},\hat{\Sigma}) \equiv nM$ (Problem 1) so that in typical instances $S(\hat{\theta},\hat{\Sigma}) \doteq nM$. If nM is larger than 100 the difference between what we recommend here and what one usually encounters in the applied literature is slight.

In notation used here, the matrix $\hat{C}$ of Chapter 1 is written as (Problem 2)

$$\hat{C} = \{\Sigma_{t=1}^{n}[(\partial/\partial\theta')f(x_t,\hat{\theta})]'\hat{\Sigma}^{-1}[(\partial/\partial\theta')f(x_t,\hat{\theta})]\}^{-1} ,$$

and

$$s^2 = S(\hat{\theta},\hat{\Sigma})/(nM - p).$$

Writing $\hat{h} = h(\hat{\theta})$ and $\hat{H} = H(\hat{\theta})$, the Wald test statistic is

$$W = \hat{h}'(\hat{H}\ \hat{C}\ \hat{H}')^{-1}\hat{h}/(q\ s^2) \quad .$$

One rejects the hypothesis

$$H\colon\ h(\theta^\circ) = 0$$

when W exceeds the upper $\alpha \times 100\%$ critical point of the F distribution with q numerator degrees of freedom and nM - p denominator degrees of freedom; that is when $W > F^{-1}(1 - \alpha;\ q,\ n - p)$.

Recall from Chapter 1 that a convenient method for computing W is to compute a vector of residuals $\hat{e}$ with typical element

$$\hat{e}_s = "y_s" - "f"("x_s",\hat{\theta}) = \hat{p}'_{(\alpha)} y_t - \hat{p}'_{(\alpha)} f(x_t,\hat{\theta}) \quad ,$$

compute a design matrix $\hat{F}$ with typical row

$$\hat{f}'_s = (\partial/\partial\theta')\ "f"("x_s",\hat{\theta}) = \hat{p}'_{(\alpha)}(\partial/\partial\theta')f(x_t,\hat{\theta}) \quad ,$$

fit the linear model

$$\hat{e} = \hat{F}\ \beta + u$$

by least squares, and test the hypothesis

$$H\colon\ \hat{H}\ \beta = \hat{h} \quad \text{against } A\colon\ \hat{H}\ \beta \neq \hat{h} \quad .$$

We illustrate.

EXAMPLE 1. (continued) We wish to test the hypothesis of homogeneity,

$$H\colon\ h(\theta^\circ) = 0 \quad \text{against } A\colon\ h(\theta^\circ) \neq 0$$

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

in the model with bivariate response function

$$f(x,\theta) = \begin{pmatrix} \ell n \dfrac{\theta_1 + \theta_2 x_1 + \theta_3 x_2 + \theta_4 x_3}{-1 + \theta_4 x_1 + \theta_7 x_2 + \theta_8 x_3} \\ \\ \ell n \dfrac{\theta_5 + \theta_3 x_1 + \theta_6 x_2 + \theta_7 x_3}{-1 + \theta_4 x_1 + \theta_7 x_2 + \theta_8 x_3} \end{pmatrix}$$

using the Wald test. To this end, the multivariate observations (y_t, x_t) are transformed to the univariate entities

$$"y_s" = p'_{(\alpha)} y_t \ , \quad "x_s" = (p'_{(\alpha)}, x'_t)'$$

which are then stored in the data set named WORKØ1 as shown in Figure 4. Using parameter values taken from Figure 3c, the entities

$$\hat{e}_s = "y_s" - "f"("x_s",\hat{\theta}) \ , \quad \hat{f}'_s = (\partial/\partial\theta')\ "f"("x_s",\hat{\theta})$$

are computed and stored in the data set named WORKØ2. We are now in a position to compute

$$W = \hat{h}'(\hat{H}\ \hat{C}\ \hat{H}')^{-1}\hat{h}/(q\ s^2)$$

by fitting the model

$$\hat{e}_s = \hat{f}'_s\ \beta + u_s$$

using least squares and testing

$$H\colon \hat{H}\ \beta = \hat{h} \quad \text{against } A\colon \hat{H}\ \beta \neq \hat{h}\ .$$

We have

Figure 4. Illustration of Wald Test Computations with Example 1.

SAS Statements:

```
DATA WORK01;  SET EXAMPLE1;
P1=3.76639099219;  P2=-3.865677509;  Y=P1*Y1+P2*Y2;  OUTPUT;
P1=0.0;  P2=3.339970820524;  Y=P1*Y1+P2*Y2;  OUTPUT;  DELETE;
DATA WORK02;  SET WORK01;
T1=-2.92458126;  T2=-1.28674630;  T3=0.81856986;  T4=0.36115784;
T5=-1.53758854;  T6=-1.04895916;  T7=0.03008670;  T8=-0.46742014;
PEAK=T1+T2*X1+T3*X2+T4*X3;  INTER=T5+T3*X1+T6*X2+T7*X3;
BASE=-1+T4*X1+T7*X2+T8*X3;
E=Y-(P1*LOG(PEAK/BASE)+P2*LOG(INTER/BASE));
DER_T1=P1/PEAK;  DER_T2=P1/PEAK*X1;  DER_T3=P1/PEAK*X2+P2/INTER*X1;
DER_T4=P1/PEAK*X3+(-P1-P2)/BASE*X1;  DER_T5=P2/INTER;
DER_T6=P2/INTER*X2;  DER_T7=P2/INTER*X3+(-P1-P2)/BASE*X2;
DER_T8=(-P1-P2)/BASE*X3;
PROC REG DATA=WORK02;
MODEL E = DER_T1 DER_T2 DER_T3 DER_T4 DER_T5 DER_T6 DER_T7 DER_T8 / NOINT;
HOMOGENE: TEST DER_T2+DER_T3+DER_T4=-0.10701860,
               DER_T3+DER_T6+DER_T7=-0.20030260,
               DER_T4+DER_T7+DER_T8=-0.07617560;
```

Output:

```
                                    SAS                                  1


DEP VARIABLE: E

                       SUM OF          MEAN
SOURCE     DF         SQUARES        SQUARE       F VALUE       PROB>F

MODEL       8   4.32010E-12   5.40012E-13         0.000       1.0000
ERROR     440       446.857      1.015584
U TOTAL   448       446.857

    ROOT MSE       1.007762      R-SQUARE        0.0000
    DEP MEAN    0.001628355      ADJ R-SQ       -0.0159
    C.V.           61888.34

NOTE: NO INTERCEPT TERM IS USED. R-SQUARE IS REDEFINED.


                     PARAMETER      STANDARD    T FOR H0:
VARIABLE   DF         ESTIMATE         ERROR  PARAMETER=0    PROB > |T|

DER_T1      1   -2.37028E-07      0.277909       -0.000        1.0000
DER_T2      1    3.17717E-07      0.226712        0.000        1.0000
DER_T3      1    5.36973E-08      0.080882        0.000        1.0000
DER_T4      1    1.64816E-08      0.030291        0.000        1.0000
DER_T5      1   -5.10589E-08      0.091930       -0.000        1.0000
DER_T6      1    7.81229E-08      0.083677        0.000        1.0000
DER_T7      1    5.65637E-10      0.036141        0.000        1.0000
DER_T8      1    2.78288E-08      0.019262        0.000        1.0000

TEST: HOMOGENE    NUMERATOR:     7.31205  DF:     3   F VALUE:     7.1998
                  DENOMINATOR:   1.01558  DF:   440   PROB >F :    0.0001
```

$$\hat{h} = \begin{pmatrix} -1.28674630 + 0.81856986 + 0.36115784 \\ 0.81856986 - 1.04895916 + 0.03008670 \\ 0.36115784 + 0.03008670 - 0.46742014 \end{pmatrix} \qquad \text{(from Figure 3c)}$$

$$= \begin{pmatrix} -0.10701860 \\ -0.20030260 \\ -0.07617560 \end{pmatrix}$$

$$\hat{H} = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \end{pmatrix}$$

$$\hat{h}'(\hat{H}\,\hat{C}\,\hat{H}')^{-1}\hat{h}/3 = 7.31205 \qquad \text{(from Figure 4)}$$

$$s^2 = 1.015584 \qquad \text{(from Figure 3c or 4)}$$

$$W = 7.1998 \qquad \text{(from Figure 4 or by division)} .$$

Since $F^{-1}(.95;\, 3,\, 440) = 2.61$ one rejects at the 5% level. The p-value is smaller than 0.001 as shown in Figure 4. []

Again following the ideas in Chapter 1, the Wald test statistic is approximately distributed as the non-central F-distribution, with q numerator degrees of freedom, nM - p denominator degrees of freedom, and non-centrality parameter

$$\lambda = h'(\theta^\circ)[H(\theta^\circ)\; C(\theta^\circ)H'(\theta^\circ)]^{-1}\; h(\theta^\circ)/2$$

$$C(\theta) = \{\Sigma_{t=1}^{n}[(\partial/\partial\theta')f(x_t,\theta)]'\; \Sigma^{-1}[(\partial/\partial\theta')f(x_t,\theta)]\}^{-1} \;;$$

written more compactly as $W \doteq F'(q, nM - p, \lambda)$. As noted in Chapter 1, the computation of λ is little different from the computation of W itself; we illustrate.

EXAMPLE 1. (continued) Consider finding the probability that a 5% level Wald test rejects the hypothesis of homogeneity

$$H: h(\theta) = \begin{pmatrix} \theta_2 + \theta_3 + \theta_4 \\ \theta_3 + \theta_6 + \theta_7 \\ \theta_4 + \theta_7 + \theta_8 \end{pmatrix} = 0$$

at the parameter settings

$$\theta^\circ = \begin{pmatrix} -2.82625314 \\ -1.25765338 \\ 0.83822896 \\ 0.36759231 \\ -1.56498719 \\ -0.98193861 \\ 0.04422702 \\ -0.44971643 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 0.16492462883510 & 0.09200572942276 \\ 0.09200572942276 & 0.08964264342294 \end{pmatrix}, \quad n = 22$$

for data with bivariate response function

$$f(x,\theta) = \begin{pmatrix} \ell n \dfrac{\theta_1 + \theta_2 x_1 + \theta_3 x_2 + \theta_4 x_3}{-1 + \theta_4 x_1 + \theta_7 x_2 + \theta_8 x_3} \\ \\ \ell n \dfrac{\theta_5 + \theta_3 x_1 + \theta_6 x_2 + \theta_7 x_3}{-1 + \theta_4 x_1 + \theta_7 x_2 + \theta_8 x_3} \end{pmatrix}$$

the value of θ° chosen is midway on the line segment joining the last two columns of Table 2.

Recall (Figure 3b) that Σ^{-1} factors as $\Sigma^{-1} = P'P$ with

$$P = \begin{pmatrix} 3.76639099219 & -3.865677509 \\ 0 & 3.339970820524 \end{pmatrix}.$$

Exactly as in Figure 4, the multivariate model is transformed in Figure 5 to a univariate model and the Jacobian of the univariate model evaluated at θ°, denote it as

Figure 5. Illustration of Wald Test Power Computations with Example 1.

SAS Statements:

```
DATA WORK01;  SET EXAMPLE1;
P1=3.76639099219;  P2=-3.865677509;  Y=P1*Y1+P2*Y2;  OUTPUT;
P1=0.0;  P2=3.339970820524;  Y=P1*Y1+P2*Y2;  OUTPUT;  DELETE;
DATA WORK02;  SET WORK01;
T1=  -2.82625314;  T2=  -1.25765338;  T3=   0.83822896;  T4=   0.36759231;
T5=  -1.56498719;  T6=  -0.98193861;  T7=   0.04422702;  T8=  -0.44971643;
PEAK=T1+T2*X1+T3*X2+T4*X3;  INTER=T5+T3*X1+T6*X2+T7*X3;
BASE=-1+T4*X1+T7*X2+T8*X3;
DER_T1=P1/PEAK;  DER_T2=P1/PEAK*X1;  DER_T3=P1/PEAK*X2+P2/INTER*X1;
DER_T4=P1/PEAK*X3+(-P1-P2)/BASE*X1;  DER_T5=P2/INTER;
DER_T6=P2/INTER*X2;  DER_T7=P2/INTER*X3+(-P1-P2)/BASE*X2;
DER_T8=(-P1-P2)/BASE*X3;
PROC MATRIX;  FETCH F DATA=WORK02(KEEP=DER_T1-DER_T8);  C=INV(F'*F);  FREE F;
FETCH T 1 DATA=WORK02(KEEP=T1-T8);
H = 0 1 1 1 0 0 0 0 / 0 0 1 0 0 1 1 0 / 0 0 0 1 0 0 1 1 ;  H0=H*T';
LAMBDA=H0'*INV(H*C*H')*H0#/2;  PRINT LAMBDA;
```

Output:

```
                                SAS                                   1


               LAMBDA          COL1

               ROW1         3.29906
```

F, is stored in the data set named WORK$\emptyset$2. Next

$$\lambda = h'[H\ C\ H']h/2 = 3.29906 \quad \text{(Figure 5)}$$

with $h = h(\theta^\circ)$, $H = (\partial/\partial\theta')h(\theta^\circ)$, and

$$C = (F'F)^{-1} = \{\Sigma_{t=1}^{n}[(\partial/\partial\theta')f(x_t,\theta^\circ)]'\Sigma^{-1}[(\partial/\partial\theta')f(x_t,\theta^\circ)]\}^{-1}$$

is computed using straightforward matrix algebra. From the Pearson-Hartley charts of the non-central F-distribution in Scheffé (1959) we obtain

$$1 - F'(2.61;\ 3,\ 440,\ 3.29906) = .55$$

as the approximation to the probability that a 5% level Wald test rejects the hypothesis of homogeneity if the true values of θ° and Σ are as above. []

A derivation of the "likelihood ratio" test of the hypothesis

$$H\colon h(\theta^\circ) = 0 \quad \text{against} \quad A\colon h(\theta^\circ) \neq 0$$

using the ideas of Chapter 1 is straightforward. Recall that $\hat\theta$ is the unconstrained minimum of $S(\theta,\hat\Sigma)$, that $\tilde\theta$ minimizes $S(\theta,\hat\Sigma)$ subject to $h(\theta) = 0$, and that $h(\theta)$ maps $\mathbb{R}^p$ into $\mathbb{R}^q$. As we have seen, an alternative method of computing $\tilde\theta$ makes use of the equivalent form of the hypothesis

$$H\colon \theta^\circ = g(\rho^\circ) \text{ for some } \rho^\circ \text{ against } A\colon \theta^\circ \neq g(\rho) \text{ for any } p\ .$$

One computes the unconstrained minimum $\hat\rho$ of $S[g(\rho),\hat\Sigma]$ and puts $\tilde\theta = g(\hat\rho)$. Using the formula given in Chapter 1,

$$L = \frac{(\text{SSE reduced} - \text{SSE full})/q}{(\text{SSE full})/(\text{"n"} - p)} ,$$

and using

$$S(\theta,\hat{\Sigma}) = \Sigma_{s=1}^{nM} [\text{"}y_s\text{"} - \text{"}f\text{"}(\text{"}x_s\text{"},\theta)]^2$$

one obtains the statistic

$$L = \frac{[S(\tilde{\theta},\hat{\Sigma}) - S(\hat{\theta},\hat{\Sigma})]/q}{S(\hat{\theta},\hat{\Sigma})/(nM - p)} .$$

One rejects H: $h(\theta^\circ) = 0$ when L exceeds the $\alpha \times 100\%$ critical point F_α of the F-distribution with q numerator degrees of freedom and nM - p denominator degrees of freedom; $F_\alpha = F^{-1}(1 - \alpha;\ q,\ nM - p)$.

We illustrate the computations with the example. In reading it, recall from Chapter 1 that one can exploit the structure of a composite function in writing code as follows. Suppose code is at hand to compute $f(x,\theta)$ and $F(x,\theta) = (\partial/\partial\theta')f(x,\theta)$. Given the value $\hat{\rho}$ compute $\tilde{\theta} = g(\hat{\rho})$ and $\tilde{G} = (\partial/\partial\theta')g(\hat{\theta})$. Obtain the value $f[x,g(\hat{\rho})]$ from the function evaluation $f(x,\tilde{\theta})$. Obtain $(\partial/\partial\rho')f[x,g(\hat{\rho})]$ by evaluating $F(x,\tilde{\theta})$ and performing the matrix multiplication $F(x,\tilde{\theta})\tilde{G}$.

EXAMPLE 1. (continued) Consider retesting the hypothesis of homogeneity, expressed as the functional dependency

$$H\colon \theta^\circ = g(\rho^\circ) \text{ for some } \rho^\circ \text{ against } A\colon \theta^\circ \neq g(\rho) \text{ for any } \rho$$

with

Figure 6. Example 1 Fitted by Multivariate Least Squares, Across Equation Constraints Imposed, Homogeneity Imposed.

SAS Statements:

```
DATA WORK01;  SET EXAMPLE1;
P1=3.76639099219;  P2=-3.865677509;  Y=P1*Y1+P2*Y2;  OUTPUT;
P1=0.0;  P2=3.339970820524;  Y=P1*Y1+P2*Y2;  OUTPUT;  DELETE;
PROC NLIN DATA=WORK01 METHOD=GAUSS ITER=50 CONVERGENCE=1.E-13;
PARMS R1=-3 R2=.8 R3=.4 R4=-1.5 R5=.03;
T1=R1; T2=-R2-R3; T3=R2; T4=R3; T5=R4; T6=-R5-R2; T7=R5; T8=-R5-R3;
PEAK=T1+T2*X1+T3*X2+T4*X3;  INTER=T5+T3*X1+T6*X2+T7*X3;
BASE=-1+T4*X1+T7*X2+T8*X3;
MODEL Y=P1*LOG(PEAK/BASE)+P2*LOG(INTER/BASE);
DER_T1=P1/PEAK;  DER_T2=P1/PEAK*X1;  DER_T3=P1/PEAK*X2+P2/INTER*X1;
DER_T4=P1/PEAK*X3+(-P1-P2)/BASE*X1;  DER_T5=P2/INTER;
DER_T6=P2/INTER*X2;  DER_T7=P2/INTER*X3+(-P1-P2)/BASE*X2;
DER_T8=(-P1-P2)/BASE*X3;
DER.R1=DER_T1; DER.R2=-DER_T2+DER_T3-DER_T6; DER.R3=-DER_T2+DER_T4-DER_T8;
DER.R4=DER_T5; DER.R5=-DER_T6+DER_T7-DER_T8;
```

Output:

```
                                   SAS                                      1

                 NON-LINEAR LEAST SQUARES ITERATIVE PHASE

          DEPENDENT VARIABLE: Y          METHOD: GAUSS-NEWTON

  ITERATION          R1              R2              R3        RESIDUAL SS
                     R4              R5

      0         -3.00000000     0.80000000      0.40000000    556.82802354
                -1.50000000     0.03000000
      .
      .
      .
      6         -2.72482606     0.85773951      0.37430609    474.68221082
                -1.59239423     0.05768367

NOTE: CONVERGENCE CRITERION MET.
                                   SAS                                      2

      NON-LINEAR LEAST SQUARES SUMMARY STATISTICS     DEPENDENT VARIABLE Y

        SOURCE                DF      SUM OF SQUARES        MEAN SQUARE

        REGRESSION             5      6441.02294156       1288.20458831
        RESIDUAL             443       474.68221082          1.07151741
        UNCORRECTED TOTAL    448      6915.70515239

        (CORRECTED TOTAL)    447       866.32697265

PARAMETER          ESTIMATE          ASYMPTOTIC            ASYMPTOTIC 95 %
                                     STD. ERROR         CONFIDENCE INTERVAL
                                                     LOWER            UPPER
R1              -2.72482606          0.17837791   -3.07540344      -2.37424867
R2               0.85773951          0.06707057    0.72592147       0.98955755
R3               0.37430609          0.02713134    0.32098315       0.42762902
R4              -1.59239423          0.07748868   -1.74468762      -1.44010083
R5               0.05768367          0.03407531   -0.00928668       0.12465402
```

$$g(\rho) = \begin{pmatrix} \rho_1 \\ -\rho_2-\rho_3 \\ \rho_2 \\ \rho_3 \\ \rho_4 \\ -\rho_5-\rho_2 \\ \rho_5 \\ -\rho_5-\rho_3 \end{pmatrix},$$

in the model with response function

$$f(x,\theta) = \begin{pmatrix} \ell n \dfrac{\theta_1 + \theta_2 x_1 + \theta_3 x_2 + \theta_4 x_3}{-1 + \theta_4 x_1 + \theta_7 x_2 + \theta_8 x_3} \\ \\ \ell n \dfrac{\theta_5 + \theta_3 x_1 + \theta_6 x_2 + \theta_7 x_3}{-1 + \theta_4 x_1 + \theta_7 x_2 + \theta_8 x_3} \end{pmatrix}$$

using the "likelihood ratio" test; θ has length $p = 8$ and ρ has length $r = 5$ whence $q = p - r = 3$. The model is bivariate so $M = 2$ and there are $n = 224$ observations. We adopt the expedient discussed immediately above, reusing the code of Figure 3c; the Jacobian of $g(\rho)$ was displayed earlier on in this section. The result is the SAS code shown in Figure 6. We obtain

$$SSE(\tilde{\theta},\hat{\Sigma}) = 474.68221082 \qquad \text{(from Figure 6)}.$$

Previously we computed

$$SSE(\hat{\theta},\hat{\Sigma}) = 446.85695247 \qquad \text{(from Figure 3c)}.$$

The "likelihood ratio" test statistic is

$$L = \frac{[S(\tilde{\theta},\hat{\Sigma}) - S(\hat{\theta},\hat{\Sigma})]/q}{S(\hat{\theta},\hat{\Sigma})/(nM - p)}$$

$$= \frac{(474.68221082 - 446.85695247)/3}{446.85695247/(448 - 8)}$$

$$= 9.133 .$$

Comparing with the critical point

$$F^{-1}(.95;\ 3,440) = 2.61$$

one rejects the hypothesis of homogeneity at the 5% level. This is, by and large, a repetition of the computations displayed in Table 3; the slight change in $\hat{\Sigma}$ has made little difference. []

In order to approximate the power of the "likelihood ratio" test we proceed as before. We formally treat the transformed model

$$"y_s" = "f"("x_s",\theta) + "e_s" \qquad s = 1, 2, \ldots, nM$$

as if it were a univariate nonlinear regression model and apply the results of Chapter 1. In a power computation, one is given an expression for the response function $f(x,\theta)$ (with range in $\mathbb{R}^M$), values for the parameters θ° and Σ, a sequence of independent variables $\{x_t\}_{t=1}^n$ and the hypothesis

$$H\colon \theta^\circ = g(\rho^\circ) \text{ for some } \rho^\circ \text{ against } A\colon \theta^\circ \neq g(\rho) \text{ for any } \rho .$$

Recall that the univariate response function is computed by factoring Σ^{-1} as $\Sigma^{-1} = P'P$ and putting

$$"f"("x_s",\theta) = p'_{(\alpha)} f(x_t,\theta)$$

for

$$s = M(t - 1) + \alpha \qquad \alpha = 1, 2, \ldots, M; \; t = 1, 2, \ldots, n \, .$$

Applying the ideas of Chapter 1, the null hypothesis induces the location parameter

$$\theta_n^* = g(\rho_n^\circ)$$

where ρ_n° is computed by minimizing

$$\Sigma_{s=1}^{nM} \{"f"("x_s",\theta^\circ) - "f"["x_s",g(\rho)]\}^2$$

$$= \Sigma_{t=1}^{n}\{f(x_t,\theta^\circ) - f[x_t,g(\rho)]\}'\Sigma_{\alpha=1}^{M}P_{(\alpha)}P'_{(\alpha)}\{f(x_t,\theta^\circ) - f[x_t,g(\rho)]\}$$

$$= \Sigma_{t=1}^{n}\{f(x_t,\theta^\circ) - f[x_t,g(\rho)]\}'\Sigma^{-1}\{f(x_t,\theta^\circ) - f[x_t,g(\rho)]\} \, .$$

Let

$$\delta_t = f(x_t,\theta^\circ) - f[x_t,g(\rho_n^\circ)]$$

$$F_t = (\partial/\partial\theta') \; f(x_t,\theta^\circ) \, .$$

Similar algebra results in the following expressions for the non-centrality parameters of Section 5 of Chapter 1

$$\lambda_1 = (\delta' P_F \delta - \delta' P_{FG} \delta)/2$$

$$\lambda_2 = (\delta'\delta - \delta' P_F \delta)/2$$

$$\delta'\delta = \Sigma_{t=1}^{n}\delta_t'\Sigma^{-1}\delta_t$$

$$\delta' P_F \delta = (\Sigma_{t-1}^{n}\delta_t'\Sigma^{-1}F_t)(\Sigma_{t=1}^{n}F_t'\Sigma^{-1}F_t)^{-1}(\Sigma_{t=1}^{n}F_t'\Sigma^{-1}\delta_t)$$

$$\delta' P_{FG} \delta = (\Sigma_{t=1}^{n}\delta_t'\Sigma^{-1}F_t G)(G'\Sigma_{t=1}^{n}F_t'\Sigma^{-1}F_t G)^{-1}(G'\Sigma_{t=1}^{n}F_t'\Sigma^{-1}\delta_t) \, .$$

One approximates the probability that the "likelihood ratio" rejects H by

$$P(L > F_\alpha) \doteq 1 - H(c_\alpha; \, q, \, nM - p, \, \lambda_1,\lambda_2)$$

where

$$c_\alpha = 1 + q\,F_\alpha/(nM - p);$$

$H(x;v_1, v_2, \lambda_1,\lambda_2)$ is the distribution defined and partially tabled in Section 5 of Chapter 1. Recall that if λ_2 is small the approximation

$$P(L > F_\alpha) \doteq 1 - F'(F_\alpha;\, q,\, nM - p,\, \lambda_1)$$

is adequate where, recall, $F'(x;\, v_1,\, v_2,\, \lambda)$ denotes the non-central F-distribution. We illustrate with the example.

EXAMPLE 1. (continued) Consider finding the probability that a 5% level "likelihood ratio" test rejects the hypothesis of homogeneity

$$H\colon\ \theta^\circ = g(\rho^\circ) \text{ for some } \rho^\circ \text{ against } A\colon\ \theta^\circ \neq g(\rho) \text{ for any } \rho$$

with

$$g(\rho) = \begin{pmatrix} \rho_1 \\ -\rho_2-\rho_3 \\ \rho_2 \\ \rho_3 \\ \rho_4 \\ -\rho_5-\rho_2 \\ \rho_5 \\ -\rho_5-\rho_3 \end{pmatrix}$$

at the parameter settings

$$\theta^\circ = \begin{pmatrix} -2.82625314 \\ -1.25765338 \\ 0.83822896 \\ 0.36759231 \\ -1.56498719 \\ -0.98193861 \\ 0.04422702 \\ -0.44971643 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 0.16492462883510 & 0.09200572942276 \\ 0.09200572942276 & 0.08964264342294 \end{pmatrix}, \quad n = 22$$

for data with bivariate response function

$$f(x,\theta) = \begin{pmatrix} \ell n \dfrac{\theta_1 + \theta_2 x_1 + \theta_3 x_2 + \theta_4 x_3}{-1 + \theta_4 x_1 + \theta_7 x_2 + \theta_8 x_3} \\ \\ \ell n \dfrac{\theta_5 + \theta_3 x_1 + \theta_6 x_2 + \theta_7 x_3}{-1 + \theta_4 x_1 + \theta_7 x_2 + \theta_8 x_3} \end{pmatrix}$$

the value of θ° chosen is midway on the line segment joining the last two columns of Table 2. Recall (Figure 3b) that Σ^{-1} factors as $\Sigma^{-1} = P'P$ with

$$P = \begin{pmatrix} 3.76639099219 & -3.865677509 \\ 0 & 3.339970820524 \end{pmatrix} .$$

Referring to Figure 7, the multivariate model is converted to a univariate model and the entities "f"("x_s",θ°) and $(\partial/\partial\theta')$"f"("$x_s$",$\theta$) are computed and stored in the data set named WORKØ2. Reusing the code of Figure 6, ρ_n° to minimize

$$\Sigma_{s=1}^{nM}\{\text{"f"("}x_s\text{"},\theta^\circ) - \text{"f"["}x_s\text{"},g(\rho)]\}^2$$

is computed using PROC NLIN. From this value and setting $\theta_n^* = g(\rho_n^\circ)$, the entities

$$\text{"}\delta_s\text{"} = \text{"f"("}x_s\text{"},\theta^\circ) - \text{"f"("}x_s\text{"},\theta_n^*)$$

$$(\partial/\partial\theta')\text{"f"("}x_s\text{"},\theta^\circ)\ (\partial/\partial\rho')g(\rho_n^\circ)$$

are computed, adjoined to the data in WORKØ2, and stored in the data set named WORKØ3. Then, as explained in connection with Figure 11a of Chapter 1, one can regress "δ_s" on $(\partial/\partial\theta')$ "f"("x_s",θ°) to obtain $\delta'\delta$ and $\delta' P_F \delta$ from the analysis of variance table and can regress "δ_s" on $(\partial/\partial\theta')$ "f"("x_s", θ°)$(\partial/\partial\rho')g(\rho_n^\circ)$ to obtain $\delta' P_{FG}\delta$. We have

Figure 7. Illustration of Likelihood Ratio Test Power Computations with Example 1.

SAS Statements:

```
DATA WORK01;  SET EXAMPLE1;
P1=3.76639099219;  P2=-3.865677509;  Y=P1*Y1+P2*Y2;  OUTPUT;
P1=0.0;  P2=3.339970820524;  Y=P1*Y1+P2*Y2;  OUTPUT;  DELETE;
DATA WORK02;  SET WORK01;
T1=  -2.82625314;  T2=  -1.25765338;  T3=   0.83822896;  T4=   0.36759231;
T5=  -1.56498719;  T6=  -0.98193861;  T7=   0.04422702;  T8=  -0.44971643;
PEAK=T1+T2*X1+T3*X2+T4*X3;  INTER=T5+T3*X1+T6*X2+T7*X3;
BASE=-1+T4*X1+T7*X2+T8*X3;
F1=P1/PEAK;  F2=P1/PEAK*X1;  F3=P1/PEAK*X2+P2/INTER*X1;
F4=P1/PEAK*X3+(-P1-P2)/BASE*X1;  F5=P2/INTER;
F6=P2/INTER*X2;  F7=P2/INTER*X3+(-P1-P2)/BASE*X2;  F8=(-P1-P2)/BASE*X3;
YDUMMY=P1*LOG(PEAK/BASE)+P2*LOG(INTER/BASE);  DROP T1-T8;
PROC NLIN DATA=WORK02 METHOD=GAUSS ITER=50 CONVERGENCE=1.E-13;
PARMS R1=-3 R2=.8 R3=.4 R4=-1.5 R5=.03;
T1=R1; T2=-R2-R3; T3=R2; T4=R3; T5=R4; T6=-R5-R2; T7=R5; T8=-R5-R3;
PEAK=T1+T2*X1+T3*X2+T4*X3;  INTER=T5+T3*X1+T6*X2+T7*X3;
BASE=-1+T4*X1+T7*X2+T8*X3;
MODEL YDUMMY=P1*LOG(PEAK/BASE)+P2*LOG(INTER/BASE);
DER_T1=P1/PEAK;  DER_T2=P1/PEAK*X1;  DER_T3=P1/PEAK*X2+P2/INTER*X1;
DER_T4=P1/PEAK*X3+(-P1-P2)/BASE*X1;  DER_T5=P2/INTER;
DER_T6=P2/INTER*X2;  DER_T7=P2/INTER*X3+(-P1-P2)/BASE*X2;
DER_T8=(-P1-P2)/BASE*X3;
DER.R1=DER_T1; DER.R2=-DER_T2+DER_T3-DER_T6; DER.R3=-DER_T2+DER_T4-DER_T8;
DER.R4=DER_T5; DER.R5=-DER_T6+DER_T7-DER_T8;
```

Output:

```
                                SAS                                       1

                 NON-LINEAR LEAST SQUARES ITERATIVE PHASE

         DEPENDENT VARIABLE: Y           METHOD: GAUSS-NEWTON

 ITERATION            R1             R2              R3          RESIDUAL SS
                      R4             R5

     0         -3.00000000     0.80000000      0.40000000      90.74281456
               -1.50000000     0.03000000

     .
     .
     .

     4         -2.73217450     0.85819875      0.37461886       7.43156658
               -1.59899262     0.05540598

NOTE: CONVERGENCE CRITERION MET.
```

Figure 7 (Continued).

SAS Statements:

```
DATA WORK03;  SET WORK02;
R1=-2.73217450;  R2=0.85819875;  R3=0.37461886;  R4=-1.59899262;  R5=0.05540598;
T1=R1; T2=-R2-R3; T3=R2; T4=R3; T5=R4; T6=-R5-R2; T7=R5; T8=-R5-R3;
PEAK=T1+T2*X1+T3*X2+T4*X3;  INTER=T5+T3*X1+T6*X2+T7*X3;
BASE=-1+T4*X1+T7*X2+T8*X3;
DELTA=P1*LOG(PEAK/BASE)+P2*LOG(INTER/BASE)-YDUMMY;
FG1=F1; FG2=-F2+F3-F6; FG3=-F2+F4-F8;FG4=F5; FG5=-F6+F7-F8;
PROC REG DATA=WORK03;  MODEL DELTA = F1-F8 / NOINT;
PROC REG DATA=WORK03;  MODEL DELTA = FG1-FG5 / NOINT;
```

Output:

SAS 3

DEP VARIABLE: DELTA

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB>F
MODEL	8	7.401094	0.925137	13358.456	0.0001
ERROR	440	0.030472	.00006925477		
U TOTAL	448	7.431567			

SAS 4

DEP VARIABLE: DELTA

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB>F
MODEL	5	0.134951	0.026990	1.639	0.1472
ERROR	443	7.296616	0.016471		
U TOTAL	448	7.431567			

$$\delta'\delta = 7.431567 \qquad \text{(from Figure 7)}$$

$$\delta' P_F \delta = 7.401094 \qquad \text{(from Figure 7)}$$

$$\delta' P_{FG} \delta = 0.134951 \qquad \text{(from Figure 7)}$$

whence

$$\begin{aligned}\lambda_1 &= (\delta' P_F \delta - \delta' P_{FG} \delta)/2 \\ &= (7.401094 - 0.134951)/2 \\ &= 3.63307\end{aligned}$$

$$\begin{aligned}\lambda_2 &= (\delta'\delta - \delta' P_F \delta)/2 \\ &= (7.431567 - 7.401094)/2 \\ &= 0.01524\end{aligned}$$

$$\begin{aligned}c_\alpha &= 1 + q\, F_\alpha/(nM - p) \\ &= 1 + 3(2.61)/(448 - 8) \\ &= 1.01780.\end{aligned}$$

Direct computation (Gallant, 1975) yields

$$\begin{aligned}P(L > 2.61) &\doteq 1 - H(1.01780;\, 3,\, 440,\, 3.63307,\, 0.01524) \\ &= 0.610\ .\end{aligned}$$

From the Pearson-Hartley charts of the non-central F-distribution (Scheffé, 1959) one has

$$\begin{aligned}P(L > 2.61) &\doteq 1 - F'(2.61;\, 3,\, 440,\, 3.63307) \\ &= 0.60\ . \quad []\end{aligned}$$

In Chapter 1 we noted that the Lagrange multiplier test had rather bizarre structural characteristics. Take the simple case of testing $H: \theta^{\circ} = \theta^{*}$ against $A: \theta^{\circ} \neq \theta^{*}$. If θ^{*} is near a local minimum or a local maximum of the sum of squares surface then the test will accept H no matter how large is the distance between $\hat{\theta}$ and θ^{*} . Also we saw some indications that the Lagrange multiplier test had poorer power than the likelihood ratio test. Thus, it would seem that one would not use the Lagrange multiplier test unless the computation of the unconstrained estimator $\hat{\theta}$ is inordinately burdensome for some reason. We shall assume that this is the case.

If $\hat{\theta}$ is inordinately burdensome to compute then $\hat{\theta}^{\#}$ will be as well. Thus, it is unreasonable to assume that one has available an estimator $\hat{\Sigma}$ with $\sqrt{n}(\hat{\Sigma} - \Sigma)$ bounded in probability when $h(\theta^{\circ}) = 0$ is false since such an estimator will almost always have to be computed from residuals from an unconstrained fit. The exception is when one has replicates at some settings of the independent variable. Accordingly, we shall base the Lagrange multiplier statistic on an estimator $\tilde{\Sigma}_n$ computed as follows.

If the hypothesis is written as a parametric restriction

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

then let $\tilde{\theta}^{\#}$ minimize $S(\theta,I)$ subject to $h(\theta) = 0$ and put

$$\tilde{\Sigma} = (1/n)\Sigma_{t=1}^{n}[y_t - f(x_t,\tilde{\theta}^{\#})][y_t - f(x_t,\tilde{\theta}^{\#})]' \ .$$

If the hypothesis is written as a functional dependency

$$H: \theta^{\circ} = g(\rho^{\circ}) \text{ for some } \rho^{\circ} \text{ against } A: \theta^{\circ} \neq g(\rho) \text{ for any } \rho$$

then let $\hat{\rho}^{\#}$ minimize $S[g(\rho),I]$ and put

$$\tilde{\Sigma} = (1/n)\ \Sigma_{t=1}^{n}\{y_t - f[x_t,g(\hat{\rho}^{\#})]\}\{y_t - f[x_t,g(\hat{\rho}^{\#})]\}' \ .$$

The constrained estimator corresponding to this estimator of scale is $\tilde{\tilde{\theta}}$ that minimizes $S(\theta,\tilde{\Sigma})$ subject to $h(\theta) = 0$. Equivalently, let $\hat{\rho}$ minimize $S[g(\rho),\tilde{\Sigma}]$ whence $\tilde{\tilde{\theta}} = g(\hat{\rho})$.

Factoring $\tilde{\Sigma}^{-1}$ as $\tilde{\Sigma}^{-1} = \tilde{P}'\tilde{P}$, denoting a typical row of $\tilde{P}$ by $\tilde{p}'_{(\alpha)}$ and formally treating the transformed model

$$"y_s" = "f"("x_s",\theta) + "e_s" \qquad s = 1, 2, \ldots, nM$$

with $s = M(t - 1) + \alpha$

$$"y_s" = \tilde{p}'_{\alpha} y_t$$

$$"x_s" = (\tilde{p}'_{(\alpha)}, x'_t)'$$

$$"f"("x_s",\theta) = \tilde{p}'_{(\alpha)} f(x_t,\theta)$$

as a univariate model, one obtains as the second version of the Lagrange multiplier test given in Chapter 1 the statistic

$$\begin{aligned}\tilde{R} = {} & [nM/S(\tilde{\tilde{\theta}},\tilde{\Sigma})]\{\Sigma_{t=1}^{n}[y_t - f(x_t,\tilde{\tilde{\theta}})]'\tilde{\Sigma}^{-1}[(\partial/\partial\theta')f(x_t,\tilde{\tilde{\theta}})]\} \\ & \times \{\Sigma_{t=1}^{n}[(\partial/\partial\theta')f(x_t,\tilde{\tilde{\theta}})]'\tilde{\Sigma}^{-1}[(\partial/\partial\theta')f(x_t,\tilde{\tilde{\theta}})]\}^{-1} \\ & \times \{\Sigma_{t=1}^{n}[(\partial/\partial\theta')f(x_t,\tilde{\tilde{\theta}})]'\tilde{\Sigma}^{-1}[y_t - f(x_t,\tilde{\tilde{\theta}})]\} .\end{aligned}$$

One rejects $H\colon h(\theta^{\circ}) = 0$ if $\tilde{R} > d_{\alpha}$ where

$$d_{\alpha} = nMF_{\alpha}/[(nM - p)/q + F_{\alpha}]$$

and F_{α} denotes the $\alpha \times (100\%)$ critical point of the F-distribution with q numerator degrees of freedom and $nM - p$ denominator degrees of freedom; that is $\alpha = 1 - F(F_{\alpha};\, q,\, nM - p)$.

One can use the same approach used in Chapter 1 to compute $\tilde{R}$. Create a data set with observations

$$"\tilde{e}_s" = "y_s" - "f"("x_s",\tilde{\tilde{\theta}}) = \tilde{p}'_{(\alpha)}y_t - p'_{(\alpha)}f(x_t,\tilde{\tilde{\theta}})$$

$$\tilde{f}'_s = (\partial/\partial\theta')"f"("x_s",\tilde{\tilde{\theta}}) = p'_{(\alpha)}(\partial/\partial\theta')f(x_t,\tilde{\tilde{\theta}}) \quad .$$

Let $\tilde{e}$ be the nM - vector with $"\tilde{e}_s"$ as elements and let $\tilde{F}$ be the nM by p matrix with $\tilde{f}'_s$ as a typical row. A linear regression of $\tilde{e}$ on $\tilde{F}$ with no intercept term yields the analysis of variance table

Source	d.f.	Sum of Squares
Regression	p	$\tilde{e}'\tilde{F}(\tilde{F}'\tilde{F})^{-1}\tilde{F}'\tilde{e}$
Error	nM-p	$\tilde{e}'\tilde{e} - \tilde{e}'\tilde{F}(\tilde{F}'\tilde{F})^{-1}\tilde{F}'\tilde{e}$
Total	nM	$\tilde{e}'\tilde{e}$.

From this table $\tilde{R}$ is computed as

$$\tilde{R} = nM\ \tilde{e}'\tilde{F}(\tilde{F}'\tilde{F})^{-1}\tilde{F}'\tilde{e}/\tilde{e}'\tilde{e} \ .$$

Let us illustrate.

EXAMPLE 1. (continued) Consider retesting the hypothesis of homogeneity, expressed as the functional dependency

$$H\colon \theta^\circ = g(\rho^\circ) \text{ for some } \rho^\circ \text{ against } A\colon \theta^\circ \neq g(\rho) \text{ for any } \rho$$

with

$$g(\rho) = \begin{pmatrix} \rho_1 \\ -\rho_2-\rho_3 \\ \rho_2 \\ \rho_3 \\ \rho_4 \\ -\rho_5-\rho_2 \\ \rho_5 \\ -\rho_5-\rho_3 \end{pmatrix}$$

in the model with response function

$$f(x,\theta) = \begin{pmatrix} \ell n \dfrac{\theta_1 + \theta_2 x_1 + \theta_3 x_2 + \theta_4 x_3}{-1 + \theta_4 x_1 + \theta_7 x_2 + \theta_8 x_3} \\ \\ \ell n \dfrac{\theta_5 + \theta_3 x_1 + \theta_6 x_2 + \theta_7 x_3}{-1 + \theta_4 x_1 + \theta_7 x_2 + \theta_8 x_3} \end{pmatrix}$$

using the Lagrange multiplier test. Note that θ is a p-vector with $p = 8$, ρ is an r-vector with $r = 5$, $q = p - r = 3$, there are M equations with $M = 2$, and n observations with $n = 224$.

Before computing the Lagrange multiplier statistic $\tilde{R}$ one must first compute $\tilde{\theta}^{\#}$ as shown in Figure 8a, $\tilde{\Sigma}$ as shown in Figure 8b, and $\tilde{\theta}$ as shown in Figure 8c. The SAS code shown in Figures 8a through 8c is simply the same code shown in Figures 3a through 3c modified by substitutions from Figure 6 so that $S[g(\rho),\Sigma]$ is minimized instead of $S(\theta,\Sigma)$. This substitution is so obvious that the discussion associated with Figures 3a, 3b, 3c, and 6 ought to suffice as a discussion on Figures 8a, 8b, 8c.

We have

$$\tilde{P} = \begin{pmatrix} 3.565728486712 & -3.75526011819 \\ 0 & 3.328902782166 \end{pmatrix} \qquad \text{(from Figure 8b)},$$

$$\hat{\rho} = \begin{pmatrix} -2.73001786 \\ 0.85800567 \\ 0.37332245 \\ -1.59315750 \\ 0.05863267 \end{pmatrix} \qquad \text{(from Figure 8c)},$$

Figure 8a. Example 1 Fitted by Least Squares, Across Equation Constraints Imposed, Homogeneity Imposed.

SAS Statements:

```
DATA WORK01;  SET EXAMPLE1;
P1=1.0;  P2=0.0;  Y=P1*Y1+P2*Y2;  OUTPUT;
P1=0.0;  P2=1.0;  Y=P1*Y1+P2*Y2;  OUTPUT;  DELETE;
PROC NLIN DATA=WORK01 METHOD=GAUSS ITER=50 CONVERGENCE=1.E-13;
PARMS R1=-3 R2=.8 R3=.4 R4=-1.5 R5=.03;
T1=R1; T2=-R2-R3; T3=R2; T4=R3; T5=R4; T6=-R5-R2; T7=R5; T8=-R5-R3;
PEAK=T1+T2*X1+T3*X2+T4*X3;  INTER=T5+T3*X1+T6*X2+T7*X3;
BASE=-1+T4*X1+T7*X2+T8*X3;
MODEL Y=P1*LOG(PEAK/BASE)+P2*LOG(INTER/BASE);
DER_T1=P1/PEAK;  DER_T2=P1/PEAK*X1;  DER_T3=P1/PEAK*X2+P2/INTER*X1;
DER_T4=P1/PEAK*X3+(-P1-P2)/BASE*X1;  DER_T5=P2/INTER;
DER_T6=P2/INTER*X2;  DER_T7=P2/INTER*X3+(-P1-P2)/BASE*X2;
DER_T8=(-P1-P2)/BASE*X3;
DER.R1=DER_T1; DER.R2=-DER_T2+DER_T3-DER_T6; DER.R3=-DER_T2+DER_T4-DER_T8;
DER.R4=DER_T5; DER.R5=-DER_T6+DER_T7-DER_T8;
OUTPUT OUT=WORK02 RESIDUAL=E;
```

Output:

```
                                   SAS                                     1

                   NON-LINEAR LEAST SQUARES ITERATIVE PHASE

               DEPENDENT VARIABLE: Y         METHOD: GAUSS-NEWTON

  ITERATION             R1             R2             R3       RESIDUAL SS
                        R4             R5

      0        -3.00000000     0.80000000     0.40000000       63.33812691
               -1.50000000     0.03000000

      .
      .
      .

      6        -2.71995278     0.80870662     0.36225861       60.25116542
               -1.53399610     0.08112412

NOTE: CONVERGENCE CRITERION MET.
```

Figure 8b. Contemporaneous Variance-Covariance Matrix of Example 1 Estimated from Least Squares Residuals, Across Equation Constraints Imposed, Homogeneity Imposed.

SAS Statements:

```
DATA WORK03;  SET WORK02;   E1=E;  IF MOD(_N_,2)=0 THEN DELETE;
DATA WORK04;  SET WORK02;   E2=E;  IF MOD(_N_,2)=1 THEN DELETE;
DATA WORK05;  MERGE WORK03 WORK04;  KEEP E1 E2;
PROC MATRIX FW=20; FETCH E DATA=WORK05(KEEP=E1 E2);
SIGMA=E'*E#/224;  PRINT SIGMA;  P=HALF(INV(SIGMA));  PRINT P;
```

Output:

```
                                        SAS                                      3


SIGMA                 COL1                COL2

ROW1        0.178738689442    0.09503630224405
ROW2      0.09503630224405    0.09023972761352


P                     COL1                COL2

ROW1        3.565728486712     -3.75526011819
ROW2                     0     3.328902782166
```

Figure 8c. Example 1 Fitted by Multivariate Least Squares, Across Equation Constraints Imposed, Homogeneity Imposed.

SAS Statements:

```
DATA WORK01;  SET EXAMPLE1;
P1=3.565728486712;  P2=-3.75526011819;  Y=P1*Y1+P2*Y2;  OUTPUT;
P1=0.0;  P2=3.328902782166.  Y=P1*Y1+P2*Y2;  OUTPUT;  DELETE;
PROC NLIN DATA=WORK01 METHOD=GAUSS ITER=50 CONVERGENCE=1.E-13;
PARMS R1=-3 R2=.8 R3=.4 R4=-1.5 R5=.03;
T1=R1; T2=-R2-R3; T3=R2; T4=R3; T5=R4; T6=-R5-R2; T7=R5; T8=-R5-R3;
PEAK=T1+T2*X1+T3*X2+T4*X3;  INTER=T5+T3*X1+T6*X2+T7*X3;
BASE=-1+T4*X1+T7*X2+T8*X3;
MODEL Y=P1*LOG(PEAK/BASE)+P2*LOG(INTER/BASE);
DER_T1=P1/PEAK;  DER_T2=P1/PEAK*X1;  DER_T3=P1/PEAK*X2+P2/INTER*X1;
DER_T4=P1/PEAK*X3+(-P1-P2)/BASE*X1;  DER_T5=P2/INTER;
DER_T6=P2/INTER*X2;  DER_T7=P2/INTER*X3+(-P1-P2)/BASE*X2;
DER_T8=(-P1-P2)/BASE*X3;
```

Output:

SAS 4

NON-LINEAR LEAST SQUARES ITERATIVE PHASE

DEPENDENT VARIABLE: Y   METHOD: GAUSS-NEWTON

ITERATION	R1 R4	R2 R5	R3	RESIDUAL SS
0	-3.00000000 -1.50000000	0.80000000 0.03000000	0.40000000	522.75679658
.				
.				
.				
6	-2.73001786 -1.59315750	0.85800567 0.05863167	0.37332245	447.09568448

NOTE: CONVERGENCE CRITERION MET.

SAS 5

NON-LINEAR LEAST SQUARES SUMMARY STATISTICS   DEPENDENT VARIABLE Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE
REGRESSION	5	5899.23816229	1179.84763246
RESIDUAL	443	447.09568448	1.00924534
UNCORRECTED TOTAL	448	6346.33384677	
(CORRECTED TOTAL)	447	806.65977490	

PARAMETER	ESTIMATE	ASYMPTOTIC STD. ERROR	ASYMPTOTIC 95 % CONFIDENCE INTERVAL LOWER	UPPER
R1	-2.73001786	0.17961271	-3.08302208	-2.37701364
R2	0.85800567	0.06701564	0.72629559	0.98971574
R3	0.37332245	0.02732102	0.31962671	0.42701818
R4	-1.59315750	0.07560672	-1.74175216	-1.44456284
R5	0.05863167	0.03396211	-0.00811621	0.12537956

$$\tilde{\tilde{\theta}} = g(\hat{\rho}) = \begin{pmatrix} \hat{\rho}_1 \\ -\hat{\rho}_2-\hat{\rho}_3 \\ \hat{\rho}_2 \\ \hat{\rho}_3 \\ \hat{\rho}_4 \\ -\hat{\rho}_5-\hat{\rho}_2 \\ \hat{\rho}_5 \\ -\hat{\rho}_5-\hat{\rho}_3 \end{pmatrix} = \begin{pmatrix} -2.7300179 \\ -1.2313281 \\ 0.8580057 \\ 0.3733224 \\ -1.5931575 \\ -0.9166373 \\ 0.0586317 \\ -0.4319541 \end{pmatrix} ,$$

and

$$S(\tilde{\tilde{\theta}},\tilde{\tilde{\Sigma}}) = 447.09568448 \qquad \text{(from Figure 8c)} .$$

As shown in Figure 9, from these values the entities

$$"\tilde{e}_s" = "y_s" - "f"("x_s",\tilde{\tilde{\theta}}) = \tilde{p}'_{(\alpha)}y_t - p'_{(\alpha)}f(x_t,\tilde{\tilde{\theta}})$$

and

$$\tilde{f}'_s = (\partial/\partial\theta')"f"("x_s",\tilde{\tilde{\theta}}) = \tilde{p}'_{(\alpha)}(\partial/\partial\theta)f(x_t,\tilde{\tilde{\theta}})$$

are computed and stored in the data set named WORKØ2 as

$$"\tilde{e}_s" = \text{ETILDE}$$

$$\tilde{f}'_s = (\text{DER\_T1, DER\_T2, \ldots, DER\_T8}) .$$

From the regression of $\tilde{e}_s$ on $\tilde{f}'_s$ we obtain

$$\tilde{e}'\tilde{F}(\tilde{F}'\tilde{F})^{-1}\tilde{F}'\tilde{e} = 24.696058 \qquad \text{(from Figure 9)}.$$

Recall that the parameter estimates shown in Figure 9 are a full Gauss-Newton step from $\tilde{\tilde{\theta}}$ to (hopefully) the minimizer of $S(\theta,\tilde{\Sigma})$. It is interesting to note that if these parameter estimates are added to the last column of Table 2 then the adjacent column is nearly reproduced as one might expect; replacing $\hat{\Sigma}$ by $\tilde{\Sigma}$ is apparently only a small perturbation.

Figure 9. Illustration of Lagrange Multiplier Test Computations with Example 1. with Example 1.

SAS Statements:

```
DATA WORK01;  SET EXAMPLE1;
P1=3.565728486712;  P2=-3.75526011819;  Y=P1*Y1+P2*Y2;  OUTPUT;
P1=0.0;  P2=3.328902782166;  Y=P1*Y1+P2*Y2;  OUTPUT;  DELETE;
DATA WORK02;  SET WORK01;
R1=-2.73001786; R2=0.85800567; R3=0.37332245; R4=-1.59315750; R5=0.05863167;
T1=R1; T2=-R2-R3; T3=R2; T4=R3; T5=R4; T6=-R5-R2; T7=R5; T8=-R5-R3;
PEAK=T1+T2*X1+T3*X2+T4*X3;  INTER=T5+T3*X1+T6*X2+T7*X3;
BASE=-1+T4*X1+T7*X2+T8*X3;
YTILDE=P1*LOG(PEAK/BASE)+P2*LOG(INTER/BASE);  ETILDE=Y-YTILDE;
DER_T1=P1/PEAK;  DER_T2=P1/PEAK*X1;  DER_T3=P1/PEAK*X2+P2/INTER*X1;
DER_T4=P1/PEAK*X3+(-P1-P2)/BASE*X1;  DER_T5=P2/INTER;
DER_T6=P2/INTER*X2;  DER_T7=P2/INTER*X3+(-P1-P2)/BASE*X2;
DER_T8=(-P1-P2)/BASE*X3;
PROC REG DATA=WORK02;  MODEL ETILDE=DER_T1-DER_T8 / NOINT;
```

Output:

SAS 1

DEP VARIABLE: ETILDE

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB>F
MODEL	8	24.696058	3.087007	3.216	0.0015
ERROR	440	422.400	0.959999		
U TOTAL	448	447.096			

ROOT MSE	0.979795	R-SQUARE	0.0552
DEP MEAN	0.001515266	ADJ R-SQ	0.0402
C.V.	64661.62		

NOTE: NO INTERCEPT TERM IS USED. R-SQUARE IS REDEFINED.

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T FOR H0: PARAMETER=0	PROB > !T!
DER_T1	1	-0.182493	0.284464	-0.642	0.5215
DER_T2	1	-0.045933	0.228965	-0.201	0.8411
DER_T3	1	-0.029078	0.075294	-0.386	0.6995
DER_T4	1	-0.012788	0.027678	-0.462	0.6443
DER_T5	1	0.055117	0.091764	0.601	0.5484
DER_T6	1	-0.125929	0.076238	-1.652	0.0993
DER_T7	1	-0.019402	0.033583	-0.578	0.5637
DER_T8	1	-0.039163	0.021008	-1.864	0.0630

From the computations above, we can compute

$$\tilde{R} = (nM)\tilde{e}'\tilde{F}(\tilde{F}'\tilde{F})^{-1}\tilde{F}'\tilde{e}/S(\tilde{\tilde{\theta}},\tilde{\Sigma})$$

$$= (448)(24.696058)/(447.09568448)$$

$$= 24.746$$

which we compare with

$$d_\alpha = (nM)F_\alpha/[(nM - p)/q + F_\alpha]$$

$$= (448)(2.61)/[440)/(3) + 2.61]$$

$$= 7.83 \ .$$

The null hypothesis of homogeneity is rejected. []

Power computations for the Lagrange multiplier test are rather onerous as seen from formulas given at the end of Section 3. The worst of it is the annoyance of having to evaluate the distribution function of a general quadratic form in normal variates rather than being able to use readily available tables. If one does not want to go to this bother then the power of the likelihood ratio test can be used as an approximation to the power of the Lagrange multiplier test.

We saw in Chapter 1 that, for univariate models, inferences based on the asymptotic theory of Chapter 3 are reasonably reliable in samples of moderate size, save in the case of the Wald test statistic, provided one takes the precaution of making degrees of freedom corrections and using tables of the F-distribution. This observation would carry over to the present situation if the matrix P with $P'P = \Sigma^{-1}$ used to rotate the model were known. It is the fact that one must use random $\hat{P}$ instead of known P that gives one pause in asserting that what is true in the univariate case is true in the multivariate case as well.

Below we report some simulations that confirm what intuition would lead one to expect. Dividing the Wald and "likelihood ratio" statistics by $S(\hat{\theta},\hat{\Sigma})/(nM-p)$ and using tables of F instead of tables of the χ^2-distribution does improve accuracy. The Wald test is unreliable. The sampling variation in $\hat{P}$ is deleterious and leads to the need for larger sample sizes before results can be trusted in the multivariate case than in the univariate case. Since $\tilde{P}$ has less sampling variation than $\hat{P}$, the null case Lagrange test probability statements are more reliable than "likelihood ratio" test probability statements. These interpretations of the simulations are subject to all the usual caveats associated with inductive inference. The details are as follows.

Table 4. Accuracy of Null Case Probability Statements.

Variable	Sample Size	Asymptotic Approximation	Monte Carlo Estimate	Monte Carlo Standard Error
P(W > F)	46	.05	.084	.0051
P(L > F)	46	.05	.067	.0046
P(R > d)	46	.05	.047	.0039
P(W' > F)	46	.05	.094	.0092
P(L' > F)	46	.05	.072	.0082
$E(s^2)$	46	1.00	1.063	.00083
P(W > F)	224	.05	.067	.0056
P(L > F)	224	.05	.045	.0046
P(R > d)	224	.05	.045	.0046

EXAMPLE 1. (continued) The simulations reported in Table 4 were computed as follows. The data in Table 1a was randomly resorted and the first n = 46 entries were used to form the variables

$$x_t = \ell n[(\text{peak price, intermediate price, base price})/\text{expenditure}]'$$

for t = 1, 2, ..., 46. For n = 224, the Table 1a was used in its entirety. At the null case parameter settings

$$\theta^\circ = \begin{pmatrix} -2.72482606 \\ -1.23204560 \\ 0.85773951 \\ 0.37430609 \\ -1.59239423 \\ -0.91542318 \\ 0.05768367 \\ -0.43198976 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 0.1649246288351 & 0.09200572942276 \\ 0.09200572942276 & 0.08964264342294 \end{pmatrix}$$

independent, normally distributed errors e_t each with mean zero and variance-covariance matrix Σ were generated and used to compute y_t according to

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

with

$$f(x,\theta) = \begin{pmatrix} \ell n \dfrac{\theta_1 + \theta_2 x_1 + \theta_3 x_2 + \theta_4 x_3}{-1 + \theta_4 x_1 + \theta_7 x_2 + \theta_8 x_3} \\ \\ \ell n \dfrac{\theta_5 + \theta_3 x_1 + \theta_6 x_2 + \theta_7 x_3}{-1 + \theta_4 x_1 + \theta_7 x_2 + \theta_8 x_3} \end{pmatrix} .$$

From each generated sample, the test statistics W, L, R discussed in this section and the statistics W′, L′ of Section 3 were computed for the hypothesis

$$H: \begin{pmatrix} \theta_2 + \theta_3 + \theta_4 \\ \theta_3 + \theta_6 + \theta_7 \\ \theta_4 + \theta_7 + \theta_8 \end{pmatrix} = 0 \; .$$

This process was replicated N times. The Monte Carlo estimate of, say, $P(L > F)$ is $\hat{p}$ equal to the number of times L exceeded F in the N Monte Carlo replicates divided by N; the reported standard error is $\sqrt{\hat{p}(1-\hat{p})/N}$. The value of F is computed as $.95 = F(F;\, 3,\, nM-p)$. $\mathcal{E}(s^2)$ is the average of $s_i^2 = S(\hat{\theta},\hat{\Sigma})/(nM-p)$ over the N Monte Carlo trials with standard error computed as $\sqrt{\frac{1}{N-1}\Sigma_{i=1}^{N}(s_i^2 - \bar{s}^2)/N}$, []

The formulas for test statistics that result from the "seemingly unrelated" notational scheme are aesthetically more appealing than the formulas presented thus far as noted earlier. Aside from aesthetics, they also serve nicely as mnemonics to the foregoing because in appearance they are just the obvious modifications of the formulas of Chapter 1 to account for the correlation structure of the errors. Verification that these formulas are correct is left as an exercise.

Recall that in the "seemingly unrelated" notational scheme we have M separate regressions of the sort studied in Chapter 1

$$y_\alpha = f_\alpha(\theta_\alpha^\circ) + e_\alpha \qquad \alpha = 1, 2, \ldots, M$$

with y_α, $f_\alpha(\theta_\alpha)$, and e_α being n-vectors. These are "stacked" into a single regressio

$$y = f(\theta^\circ) + e$$

by writing

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{pmatrix}_{nM \times 1}$$

$$f(\theta) = \begin{pmatrix} f_1(\theta_1) \\ f_2(\theta_2) \\ \vdots \\ f_M(\theta_M) \end{pmatrix}_{nM \times 1}$$

$$e = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_M \end{pmatrix}_{nM \times 1}$$

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_M \end{pmatrix}_{p \times 1}$$

with

$$\mathcal{E}(e) = 0 \qquad \mathcal{C}(e,e') = \Sigma \otimes I.$$

We have available some estimator $\hat{\Sigma}$ of Σ, typically that obtained by finding $\hat{\theta}^{\#}$ to minimize

$$S(\theta,\Sigma) = [y - f(\theta)]'(\Sigma^{-1} \otimes I)[y - f(\theta)]$$

with $\Sigma = I$ and taking as the estimate the matrix $\hat{\Sigma}$ with typical element

$$\hat{\sigma}_{\alpha\beta} = (1/n)[y_\alpha - f_\alpha(\hat{\theta}^\#_\alpha)]'[y_\beta - f_\beta(\hat{\theta}^\#_\beta)].$$

The estimator $\hat{\theta}$ minimizes $S(\theta,\hat{\Sigma})$. Recall that the task at hand is to test a hypothesis that can be expressed either as a parametric restriction

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

or as a functional dependency

$$H\colon \theta^\circ = g(\rho^\circ) \text{ for some } \rho^\circ \text{ against } A\colon \theta^\circ \neq g(\rho) \text{ for any } \rho$$

where ρ is an r-vector, $h(\theta)$ is a q-vector, and $p = r + q$. The various Jacobians required are

$$H(\theta) = (\partial/\partial\theta')h(\theta)$$

$$G(\rho) = (\partial/\partial\rho')g(\rho)$$

$$F(\theta) = (\partial/\partial\theta')f(\theta)$$

being q by p, p by r, and nM by p respectively.

The Wald test statistic is

$$W = \hat{h}'(\hat{H}\ \hat{C}\ \hat{H})^{-1}\hat{h}/(q\ s^2)$$

with

$$\hat{C} = [F'(\hat{\theta})(\hat{\Sigma}^{-1} \otimes I)F(\hat{\theta})]^{-1}$$

$$s^2 = S(\hat{\theta},\hat{\Sigma})/(nM - p)$$

$$\hat{h} = h(\hat{\theta}),\ \hat{H} = H(\hat{\theta}).$$

One rejects $H\colon h(\theta^\circ) = 0$ when W exceeds $F^{-1}(1-\alpha,\ q,\ nM-p)$.

The form of the "likelihood ratio" test is unaltered

$$L = \frac{[S(\tilde{\theta},\hat{\Sigma}) - S(\hat{\theta},\hat{\Sigma})]/q}{S(\hat{\theta},\hat{\Sigma})/(nM - p)}$$

where $\tilde{\theta} = g(\hat{\rho})$ and $\hat{\rho}$ minimizes $S[g(\rho),\hat{\Sigma}]$. One rejects when L exceeds $F^{-1}(1-\alpha;\ q,\ nM-p)$.

As noted above one is unlikely to use the Lagrange multiplier test unless

$S(\theta,\Sigma)$ is difficult to minimize while minimization of $S[g(\rho),\Sigma]$ is relatively easy. In this instance one is apt to use the estimate $\tilde{\Sigma}$ with typical element

$$\tilde{\sigma}_{\alpha\beta} = (1/n)[y_\alpha - f_\alpha(\tilde{\theta}^{\#}_\alpha)]'[y_\beta - f_\beta(\tilde{\theta}^{\#}_\beta)]$$

where $\tilde{\theta}^{\#} = g(\hat{\rho}^{\#})$ and $\hat{\rho}^{\#}$ minimizes $S[g(\rho), I]$. Let $\tilde{\tilde{\theta}} = g(\hat{\tilde{\rho}})$ where $\hat{\tilde{\rho}}$ minimizes $S[g(\rho),\tilde{\Sigma})$. The Gauss-Newton step away from $\tilde{\tilde{\theta}}$ (presumably) toward $\hat{\theta}$ is

$$\tilde{D} = [\tilde{F}'(\tilde{\Sigma}^{-1} \otimes I)\tilde{F}]^{-1}\, \tilde{F}'(\tilde{\Sigma}^{-1} \otimes I)[y - \tilde{f}]$$

where $\tilde{F} = F(\tilde{\tilde{\theta}})$, and $\tilde{f} = f(\tilde{\tilde{\theta}})$. The Lagrange multiplier test statistic is

$$\tilde{R} = nM\, \tilde{D}'[\tilde{F}'(\tilde{\Sigma}^{-1} \otimes I)\tilde{F}]^{-1}\, \tilde{D}/S(\tilde{\tilde{\theta}},\tilde{\Sigma}) .$$

One rejects when $\tilde{R}$ exceeds

$$d_\alpha = nM\, F_\alpha/[(nM-p)/q + F_\alpha]$$

with $F_\alpha = F^{-1}(1-\alpha;\ q,\ nM-p)$.

PROBLEMS

1. Show that if $\hat{\theta}^{\#}$ minimizes $S(\theta,I)$ and $\hat{\Sigma} = (1/n)\Sigma_{t=1}^{n}[y_t - f(x_t,\hat{\theta}^{\#})][y_t - f(x_t,\hat{\theta}^{\#})]'$ then $S(\hat{\theta},\hat{\Sigma}) = nM$.

2. Show that the matrix $\hat{C}$ of Chapter 1 can be written as

$$\hat{C} = [\Sigma_{s=1}^{nM}(\partial/\partial\theta)\text{"}f\text{"}(\text{"}x_s\text{"},\hat{\theta})\ (\partial/\partial\theta')\text{"}f\text{"}(\text{"}x_s\text{"},\hat{\theta})]^{-1}$$

using the notation of Section 2. Show that $(\partial/\partial\theta')\text{"}f\text{"}(\text{"}x_s\text{"},\hat{\theta}) = p_{(\alpha)}'(\partial/\partial\theta')f(x_t,\hat{\theta})$ whence

$$\hat{C} = \{\Sigma_{t=1}^{n}[(\partial/\partial\theta')f(x_t,\hat{\theta})]'\hat{\Sigma}^{-1}[(\partial/\partial\theta')f(x_t,\hat{\theta})]\}^{-1} \quad .$$

3. Show that the equation $s = M(t - 1) + \alpha$ uniquely defines t and α as a function of s provided that $1 \leq \alpha \leq M$ and s, t, α are positive integers.

4. Verify that the formulas given for W and $\tilde{R}$ at the end of this section in the "seemingly unrelated" notational scheme agree with the formulas that precede them.

5. CONFIDENCE INTERVALS

As discussed in Section 6 of Chapter 1, a confidence interval on any (twice continuously differentiable) parametric function $\gamma(\theta)$ can be obtained by inverting any one of the tests of

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

that were discussed in the previous section. Letting

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

one puts in the interval all those γ° for which the hypothesis $H: h(\theta^\circ) = 0$ is accepted at the α level of significance. The same approach applies to confidence regions, the only difference is that $\gamma(\theta)$ and γ° will be q-vectors instead of being univariate.

There is really nothing to add to the discussion in Section 6 of Chapter 3. The methods discussed there transfer directly to multivariate nonlinear regression. The only difference is that the test statistics W, L, and $\tilde{R}$ are computed according to the formulas of the previous section. The rest is the same.

6. MAXIMUM LIKELIHOOD ESTIMATION

Given some estimator of scale $\hat{\Sigma}_0$, the corresponding least squares estimator $\hat{\theta}_0$ minimizes $S(\theta,\hat{\Sigma}_0)$ where, recall,

$$S(\theta,\Sigma) = (1/n)\Sigma_{t=1}^{n}[y_t - f(x_t,\theta)]'\Sigma^{-1}[y_t - f(x_t,\theta)].$$

A natural tendency is to iterate by putting

$$\hat{\Sigma}_{i+1} = (1/n)\Sigma_{t=1}^{n}[y_t - f(x_t,\hat{\theta}_i)][y_t - f(x_t,\hat{\theta}_i)]',$$

and

$$\hat{\theta}_{i+1} = \operatorname{argmin}_\theta\ S(\theta,\hat{\Sigma}_{i+1})$$

where $\operatorname{argmin}_\theta\ S(\theta,\Sigma)$ means that value of θ which minimizes $S(\theta,\Sigma)$. Continuing this process generates a sequence of estimators

$$\hat{\Sigma}_0 \to \hat{\theta}_0 \to \hat{\Sigma}_1 \to \hat{\theta}_1 \to \hat{\Sigma}_2 \to \hat{\theta}_2 \to \ldots \quad .$$

If the sequence is terminated at any finite step I then $\hat{\Sigma}_I$ is a consistent estimator of scale with $\sqrt{n}\ (\hat{\Sigma}_I - \Sigma_n^\circ)$ bounded in probability under the regularity conditions listed in Section 3 (Problem 1). Thus, $\hat{\theta}_I$ is just a least squares estimator and the theory and methods discussed in Sections 1 through 5 apply. If one iterates until the sequence $\{(\hat{\theta}_i,\hat{\Sigma}_i)\}_{i=1}^{\infty}$ converges then the limits

$$\hat{\Sigma}_\infty = \ell im_{i\to\infty}\ \hat{\Sigma}_i$$

$$\hat{\theta}_\infty = \ell im_{i\to\infty}\ \hat{\theta}_i$$

will be a local maximum of a normal-errors likelihood surface provided that regularity conditions similar to those listed in Problem 4, Section 4, Chapter 1 are imposed. To see intuitively that this claim is correct, observe that under a normality assumption the random variables $\{y_t\}_{t=1}^{n}$ are independent each with density

$$n[y_t|f(x_t,\theta),\Sigma] = (2\pi)^{-M/2}(\det\ \Sigma)^{-\frac{1}{2}}\ e^{-\frac{1}{2}[y_t\ -\ f(x_t,\theta)]'\Sigma^{-1}[y_t\ -\ f(x_t,\theta)]}\ .$$

The log likelihood is

$$\ell n\ \Pi_{t=1}^{n}[y_t|f(x_t,\theta),\Sigma] = \text{const.}\ -\Sigma_{t=1}^{n}\ \tfrac{1}{2}\{\ell n\ \det\ \Sigma + [y_t\ -\ f(x_t,\theta)]'\Sigma^{-1}[y_t - f(x_t,\theta)]$$

so the maximum likelihood estimator can be characterized as that value of (θ,Σ) which minimizes

$$\begin{aligned} s_n(\theta,\Sigma) &= (1/n)\ \Sigma_{t=1}^{n}\ \tfrac{1}{2}\{\ell n\ \det\ \Sigma + [y_t\ -\ f(x_t,\theta)]'\Sigma^{-1}[y_t\ -\ f(x_t,\theta)]\} \\ &= \tfrac{1}{2}[\ell n\ \det\ \Sigma + (1/n)\ S(\theta,\Sigma)]. \end{aligned}$$

Further, $s_n(\theta,\Sigma)$ will have a local minimum at each local maximum of the likelihood surface, and conversely. By Problem 11 of Section 3 we have that

$$s_n(\hat{\theta}_i,\hat{\Sigma}_{i+1}) < s_n(\hat{\theta}_i,\hat{\Sigma}_i)$$

provided that $\hat{\Sigma}_{i+1} \neq \hat{\Sigma}_i$. By definition $S(\hat{\theta}_{i+1},\hat{\Sigma}_{i+1}) \leq S(\hat{\theta}_i,\hat{\Sigma}_{i+1})$ provided $\hat{\theta}_{i+1} \neq \hat{\theta}_i$. Arguments similar to those of Problem 4, Section 4, Chapter 1, can be employed to strengthen the weak inequality to a strict inequality. Thus we have

$$s_n(\hat{\theta}_{i+1},\hat{\Sigma}_{i+1}) < s_n(\hat{\theta}_i,\hat{\Sigma}_i)$$

unless $(\hat{\theta}_{i+1},\hat{\Sigma}_{i+1}) = (\hat{\theta}_i,\hat{\Sigma}_i)$ and can conclude that $(\hat{\theta}_{i+1},\hat{\Sigma}_{i+1})$ is downhill from $(\hat{\theta}_i,\hat{\Sigma}_i)$. By attending to a few extra details, one can conclude that the limit $(\hat{\theta}_\infty,\hat{\Sigma}_\infty)$ must exist and be a local minimum of $s_n(\theta,\Sigma)$.

One can set forth regularity conditions such that the uniform almost sure limit of $s_n(\theta,\Sigma)$ exists and has a unique minimum at (θ^*,Σ^*). This fact coupled with the fact that $(\hat{\theta}_o,\hat{\Sigma}_o)$ has almost sure limit (θ^*,Σ^*) under the regularity conditions listed in Section 3 is enough to conclude that $(\hat{\theta}_\infty,\hat{\Sigma}_\infty)$ is tail equivalent to the maximum likelihood estimator and thus for any theoretical purpose can be regarded as if it were the maximum likelihood estimator. As a practical matter one may prefer some other algorithm to iterated least squares as a means to compute the maximum likelihood estimator. In a direct computation, the number of arguments of the objective function that must be minimized can be reduced by "concentrating" the likelihood as follows. Let

$$\hat{\Sigma}(\theta) = (1/n)\Sigma_{t=1}^{n} [y_t - f(x_t,\theta)][y_t - f(x_t,\theta)]'$$

and observe that by Problems 8 and 11 of Section 3

$$\min_{\Sigma} s_n(\theta,\Sigma) = s_n[\theta,\hat{\Sigma}(\theta)] = \tfrac{1}{2}[\ell n \det \hat{\Sigma}(\theta) + M] .$$

Thus it suffices to compute

$$\hat{\theta}_\infty = \operatorname{argmin}_\theta \ell n \det \hat{\Sigma}(\theta)$$

and put

$$\hat{\Sigma}_\infty = \hat{\Sigma}(\hat{\theta}_\infty)$$

to have the minimizer $(\hat{\theta}_\infty,\hat{\Sigma}_\infty)$ of $s_n(\theta,\Sigma)$. As before, the reader is referred to Gill, Murray, and Wright (1981) for guidance in the choice of algorithms for minimizing $\ell n \det \hat{\Sigma}(\theta)$.

It seems unnecessary to set forth regularity conditions from which the claims above can be derived rigorously for two reasons. First, the intuition behind them is fairly compelling. Secondly, maximum likelihood estimation of the parameters of a multivariate nonlinear regression model is a special case of maximum likelihood estimation of the parameters of a nonlinear simultaneous equation. The general theory is treated in detail in Chapter 8 and when the general regularity conditions are specialized to the present instance the result is a listing of regularity conditions that does not differ in any essential respect from those listed in Section 3. For these two reasons it seems pointless to bother with the details.

The following facts hold under the regularity conditions listed in Chapter 8:

$\hat{\Sigma}_\infty$ is consistent for Σ^*, $\sqrt{n}\,(\hat{\Sigma}_\infty - \Sigma_n^\circ)$ is bounded in probability, and $\hat{\theta}_\infty$ minimizes $S(\theta,\hat{\Sigma}_\infty)$. It follows that $\hat{\theta}_\infty$ is a least squares estimator so that one can apply the theory and methods of Section 4 to have a methodology for inference regarding θ using maximum likelihood estimates. We shall have more to say on this later. However, for joint inference regarding (θ,Σ) or marginal inference regarding Σ one needs the joint asymptotics of $(\hat{\theta}_\infty,\hat{\Sigma}_\infty)$. This is provided by specializing the results of Chapter 3 to the present instance.

In order to develop an asymptotic theory suited to inference regarding Σ it is necessary to subject Σ_n° to a Pitman drift and thus it is necessary to use a slightly different setup for the data generating model than that used in Section 3 of this chapter. For this, we need some additional notation regarding Σ. Let σ, a vector of length $M(M+1)/2$, denote the upper triangle of Σ arranged as follows

$$\sigma = (\sigma_{11},\sigma_{12},\sigma_{22},\sigma_{13},\sigma_{23},\sigma_{33},\ldots,\sigma_{1M},\sigma_{2M},\ldots,\sigma_{MM})'$$

The mapping of σ into the elements of Σ is denoted as $\Sigma(\sigma)$. Let vec Σ denote the M^2-vector obtained by stacking the columns of $\Sigma = [\Sigma_{(1)}, \Sigma_{(2)}, \ldots, \Sigma_{(M)}]$ according to

$$\operatorname{vec} \Sigma = \begin{pmatrix} \Sigma_{(1)} \\ \Sigma_{(2)} \\ \vdots \\ \Sigma_{(M)} \end{pmatrix} .$$

The mapping of σ into vec $\Sigma(\sigma)$ is a linear map and can be written as

$$\operatorname{vec} \Sigma(\sigma) = K\,\sigma$$

where K is an M^2 by $M(M+1)/2$ matrix of zeroes and ones. Perhaps it is best to illustrate these notations with a 3 by 3 example:

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} ,$$

$$\sigma = (\sigma_{11}, \sigma_{12}, \sigma_{22}, \sigma_{13}, \sigma_{23}, \sigma_{33})' ,$$

$$\Sigma(\sigma) = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{pmatrix} ,$$

$$\underbrace{\begin{pmatrix} \sigma_{11} \\ \sigma_{12} \\ \sigma_{13} \\ \sigma_{12} \\ \sigma_{22} \\ \sigma_{23} \\ \sigma_{13} \\ \sigma_{23} \\ \sigma_{33} \end{pmatrix}}_{\text{vec } \Sigma(\sigma)} = \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \sigma_{11} \\ \sigma_{12} \\ \sigma_{22} \\ \sigma_{13} \\ \sigma_{23} \\ \sigma_{33} \end{pmatrix}}_{K\sigma} .$$

The notation $\Sigma^{\frac{1}{2}}$ denotes a matrix such that $\Sigma = (\Sigma^{\frac{1}{2}})(\Sigma^{\frac{1}{2}})'$ and the notation $\Sigma^{-\frac{1}{2}}$ denotes a matrix such that $\Sigma^{-1} = (\Sigma^{-\frac{1}{2}})'(\Sigma^{-\frac{1}{2}})$. We shall always assume that the factorization algorithm used to compute $\Sigma^{\frac{1}{2}}$ and $\Sigma^{-\frac{1}{2}}$ satisfies $\Sigma^{\frac{1}{2}}\, \Sigma^{-\frac{1}{2}} = I$.

The data generating model is

$$y_t = f(x_t, \theta_n^\circ) + [\Sigma(\sigma_n^\circ)]^{\frac{1}{2}}\, e_t$$

with θ_n° known to lie in some compact set Θ^* and σ_n° known to lie in some compact set $\mathcal{S}^*$ over which $\Sigma(\sigma)$ is a positive definite matrix, see Section 3 for a construction of such an $\mathcal{S}^*$. The functional form of $f(x,\theta)$ is known, x is k-dimensional, θ is p-dimensional, and $f(x,\theta)$ takes its values in $\mathbb{R}^M$; y_t and e_t are M-vectors. The errors e_t are independently and identically distributed each with mean zero and variance-covariance matrix the identity matrix of order M. Note that normality is not assumed in deriving the asymptotics. The parameter to be estimated is

$$\lambda_n^\circ = (\theta_n^\circ, \sigma_n^\circ) .$$

Drift is imposed so that

NOTATION 5

<u>General (Chapter 3)</u>	<u>Specific (Chapter 6)</u>
$e_t = q(y_t,x_t,\gamma_n^o)$	$e_t = [\Sigma(\sigma_n^o)]^{-\frac{1}{2}}\,[y_t - f(x_t,\theta_n^o)]$
$\gamma \in \Gamma^*$	$\gamma = (\theta',\sigma')',\ \theta \in \Theta^*,\ \sigma \in \mathfrak{S}^*$
$y = Y(e,x,\gamma)$	$y = f(x,\theta) + [\Sigma(\sigma)]^{\frac{1}{2}}e$
$s(y,x,\hat{\tau}_n,\lambda)$	$s(y,x,\lambda) = \frac{1}{2}\,\ell n \det \Sigma(\sigma) + \frac{1}{2}[y - f(x,\theta)]'\Sigma^{-1}(\sigma)[y - f(x,\theta)]$
$\hat{\tau}_n \in T$	No preliminary estimator
$\lambda \in \Lambda^*$	$\lambda = (\theta',\sigma')',\ \theta \in \Theta^*,\ \sigma \in \mathfrak{S}^*$
$s_n(\lambda) = (1/n)\Sigma_{t=1}^n s(y_t,x_t,\hat{\tau}_n,\lambda)$	$s_n(\lambda) = \frac{1}{2}\,\ell n \det \Sigma(\sigma) + (1/n)\Sigma_{t=1}^n \frac{1}{2}[y_t - f(x_t,\theta)]'\Sigma^{-1}(\sigma)[y_t - f(x_t,\theta)]$
$s_n^o(\lambda) = (1/n)\Sigma_{t=1}^n \int_{\mathcal{E}} s[Y(e,x_t,\gamma_n^o),x_t,\tau_n^o,\lambda]dP(e)$	$s_n^o(\lambda) = \frac{1}{2}\,\ell n \det \Sigma(\sigma_n^o) + \frac{1}{2}M + (1/n)\Sigma_{t=1}^n \frac{1}{2}[f(x_t,\theta_n^o) - f(x_t,\theta)]'\Sigma^{-1}(\sigma_n^o)[f(x_t,\theta_n^o) - f(x_t,$
$s^*(\lambda) = \int_{\mathcal{X}}\int_{\mathcal{E}} s[Y(e,x,\gamma^*),x,\tau^*,\lambda]dP(e)\,d\mu(x)$	$s^*(\lambda) = \frac{1}{2}\,\ell n \det \Sigma(\sigma) + \frac{1}{2}M + \int_{\mathcal{X}} \frac{1}{2}[f(x,\theta^*) - f(x,\theta)]'\Sigma^{-1}(\sigma^*)[f(x,\theta^*) - f(x,\theta)]d\mu(x)$
$\hat{\lambda}_n$ minimizes $s_n(\lambda)$	$\hat{\lambda}_n$ minimizes $s_n(\lambda)$
$\tilde{\lambda}_n$ minimizes $s_n(\lambda)$ subject to $h(\lambda) = 0$	$\tilde{\lambda}_n = g(\hat{\rho}_n)$ minimizes $s_n(\lambda)$ subject to $h(\lambda) = 0$
λ_n^o minimizes $s_n^o(\lambda)$	λ_n^o minimizes $s_n^o(\lambda)$
λ_n^* minimizes $s_n^o(\lambda)$ subject to $h(\lambda) = 0$	$\lambda_n^* = g(\rho_n^o)$ minimizes $s_n^o(\lambda)$ subject to $h(\lambda) = 0$
λ^* minimizes $s^*(\lambda)$	λ^* minimizes $s^*(\lambda)$

In order to use the formulas for the parameters of the asymptotic distributions set forth in Chapter 3 it is necessary to compute $(\partial/\partial\lambda)s[Y(e,x,\gamma^\circ),x,\lambda]$ and $(\partial^2/\partial\lambda\partial\lambda')s[Y(e,x,\gamma^\circ),x,\lambda]$. To this end, write

$$s[Y(e,x,\gamma^\circ),x,\lambda]$$

$$= \tfrac{1}{2}\,\ell n\ \det\ \Sigma(\sigma) + \tfrac{1}{2}[u + \delta(x,\theta)]'\Sigma^{-1}(\sigma)[u + \delta(x,\theta)]$$

where $u = \Sigma^{\frac{1}{2}}(\sigma^\circ)e$ and $\delta(x,\theta) = f(x,\theta^\circ) - f(x,\theta)$. Note that u has mean zero and variance-covariance matrix Σ_n°. Letting ξ_i denote a vector with a one in the i^{th} position and zeroes elsewhere, we have (Problem 2):

$(\partial/\partial\theta_i)s[Y(e,x,\gamma^\circ),x,\lambda]$

$$= -[u + \delta(x,\theta)]' \Sigma^{-1}(\sigma)\ (\partial/\partial\theta_i)f(x,\theta)$$

$(\partial/\partial\sigma_i)s[Y(e,x,\gamma^\circ),x,\lambda]$

$$= \tfrac{1}{2}\ \mathrm{tr}(\Sigma^{-1}(\sigma)\ \Sigma(\xi_i)\{I - \Sigma^{-1}(\sigma)[u - \delta(x,\theta)][u - \delta(x,\theta)]'\})$$

$(\partial^2/\partial\theta_i\partial\theta_j)s[Y(e,x,\gamma^\circ),x,\lambda]$

$$= [(\partial/\partial\theta_i)f(x,\theta)]'\Sigma^{-1}(\sigma)[(\partial/\partial\theta_j)f(x,\theta)]$$

$$- \Sigma_{\alpha=1}^{M}\ \Sigma_{\beta=1}^{M}[\xi_\alpha'\ \Sigma^{-1}(\sigma)\xi_\beta][u_\alpha + \delta_\alpha(x,\theta)](\partial^2/\partial\theta_i\partial\theta_j)f_\beta(x,\theta)$$

$(\partial^2/\partial\sigma_i\partial\theta_j)s[Y(e,x,\gamma^\circ),x,\lambda]$

$$= [u + \delta(x,\theta)]' \Sigma^{-1}(\sigma)\ \Sigma(\xi_i)\ \Sigma^{-1}(\sigma)\ (\partial/\partial\theta_i)f(x,\theta)$$

$(\partial^2/\partial\sigma_i\partial\sigma_j)s[Y(e,x,\gamma^\circ),x,\lambda]$

$$= -\tfrac{1}{2}\ \mathrm{tr}(\Sigma^{-1}(\sigma)\ \Sigma(\xi_j)\ \Sigma^{-1}(\sigma)\ \Sigma(\xi_i)\ \{I - 2\ \Sigma^{-1}(\sigma)[u - \delta(x,\theta)][u-\delta(x,\theta)]'\})$$

In order to write $(\partial/\partial\sigma)\, s[Y(e,x,\gamma^\circ),x,\lambda]$ as a vector we use the fact (Problem 3) that for conformable matrices A, B, C

$$\operatorname{vec}(ABC) = (C' \otimes A) \operatorname{vec} B$$

$$\operatorname{tr}(ABC) = (\operatorname{vec} A')'(I \otimes B) \operatorname{vec} C$$

where, recall, vec A denotes the columns of A stacked into a column vector as defined and illustrated a few paragraphs earlier and $A \otimes B$ denotes the matrix with typical block $a_{ij}B$ as defined and illustrated in Section 2 of this chapter. Recalling that $\operatorname{vec} \Sigma(\sigma) = K\,\sigma$ we have

$$\begin{aligned}
&(\partial/\partial\sigma_i) s[Y(e,x,\gamma^\circ),x,\lambda] \\
&\quad = \tfrac{1}{2} \operatorname{tr} (\Sigma^{-1}(\sigma)\ \Sigma(\xi_i)\ \Sigma^{-1}(\sigma)\{\Sigma(\sigma) - [u - \delta(x,\theta)][u - \delta(x,\theta)]'\} \\
&\quad = \tfrac{1}{2} \operatorname{vec}'[\Sigma(\xi_i)\Sigma^{-1}(\sigma)][I \otimes \Sigma^{-1}(\sigma)] \operatorname{vec} \{\Sigma(\sigma) - [u - \delta(x,\theta)][u - \delta(x,\theta)]'\} \\
&\quad = \tfrac{1}{2} \operatorname{vec}'\Sigma(\xi_i)\ [\Sigma^{-1}(\sigma) \otimes I][I \otimes \Sigma^{-1}(\sigma)] \operatorname{vec} \{\Sigma(\sigma) - [u - \delta(x,\theta)][u - \delta(x,\theta)]' \\
&\quad = \tfrac{1}{2}\, \xi_i'\ K'[\Sigma^{-1}(\sigma) \otimes \Sigma^{-1}(\sigma)] \operatorname{vec} \{\Sigma(\sigma) - [u - \delta(x,\theta)][u - \delta(x,\theta)]'\}
\end{aligned}$$

From this expression we deduce that

$$(\partial/\partial\lambda) s[Y(e,x,\gamma_n^\circ),x,\lambda_n^\circ] = \begin{pmatrix} [(\partial/\partial\theta')f(x,\theta)]'(\Sigma_n^\circ)^{-1}u \\ \tfrac{1}{2} K'(\Sigma_n^\circ \otimes \Sigma_n^\circ)^{-1} \operatorname{vec} [uu' - (\Sigma_n^\circ)^{-1}] \end{pmatrix} .$$

In terms of the notation

NOTATION 6.

$$\Omega = \int_{\mathcal{X}} [(\partial/\partial\theta')f(x,\theta^*)]'\Sigma^{-1}(\sigma^*)[(\partial/\partial\theta')f(x,\theta^*)]d\mu(x)$$

$$\mathfrak{J} = \int_{\mathcal{X}} (\partial/\partial\theta')f(x,\theta^*)d\mu(x)$$

$$\Omega_n^\circ = (1/n)\Sigma_{t=1}^n[(\partial/\partial\theta')f(x_t,\theta_n^\circ)]'\Sigma^{-1}(\sigma_n^\circ)[(\partial/\partial\theta')f(x_t,\theta_n^\circ)]$$

$$\mathfrak{J}_n^\circ = (1/n)\Sigma_{t=1}^n(\partial/\partial\theta')f(x,\theta_n^\circ)$$

we have

$$\mathcal{U}_n^\circ = 0$$

$$\mathcal{J}_n^\circ = \begin{pmatrix} \Omega_n^\circ & (\mathcal{F}_n^\circ)'(\Sigma_n^\circ)^{-1}\mathcal{E}[u\ \text{vec}'(uu')](\Sigma_n^\circ \otimes \Sigma_n^\circ)^{-1}K \\ \text{sym} & \tfrac{1}{4}K'(\Sigma_n^\circ \otimes \Sigma_n^\circ)^{-1}\text{Var}[\text{vec}(uu')](\Sigma_n^\circ \otimes \Sigma_n^\circ)^{-1}K \end{pmatrix} .$$

The third moment of a normally distributed random variable is zero whence $\mathcal{E}[u\ \text{vec}'(uu')] = 0$ under normality. From Henderson and Searle (1979) we have that under normality

$$\text{Var}[\text{vec}(uu')] = (\Sigma_n^\circ \otimes \Sigma_n^\circ)(I + I_{(M,M)})$$

where $I_{(M,M)}$ is a matrix whose entries are zeros and ones defined for a p by q matrix A as

$$\text{vec}\ A = I_{(p,q)}\ \text{vec}(A') .$$

Since for any σ

$$K\sigma = \text{vec}\ \Sigma = I_{(M,M)}\text{vec}\ \Sigma' = I_{(M,M)}\text{vec}\ \Sigma = I_{(M,M)}K\sigma$$

we must have $K = I_{(M,M)}K$ whence, under normality,

$$\mathcal{J}_n^\circ = \begin{pmatrix} \Omega_n^\circ & 0 \\ 0 & \tfrac{1}{2}K'(\Sigma_n^\circ \otimes \Sigma_n^\circ)^{-1}K \end{pmatrix} .$$

Using

$$\text{vec}(uu') = (\Sigma_n^\circ \otimes \Sigma_n^\circ)^{\frac{1}{2}}\ \text{vec}\ ee' ,$$

we have that

$$\mathcal{J}_n^\circ = \begin{pmatrix} \Omega_n^\circ & (\mathcal{F}_n^\circ)'[(\Sigma_n^{\circ\,-\frac{1}{2}})]'\mathcal{E}[e\ \text{vec}'(ee')](\Sigma_n^\circ \otimes \Sigma_n^\circ)^{-\frac{1}{2}}K \\ \text{sym} & \tfrac{1}{4}K'[(\Sigma_n^\circ \otimes \Sigma_n^\circ)^{-\frac{1}{2}}]'\ \text{Var}[\text{vec}(ee')](\Sigma_n^\circ \otimes \Sigma_n^\circ)^{-\frac{1}{2}}K \end{pmatrix}$$

in general.

The form of $\mathcal{E}(\partial^2/\partial\sigma\partial\sigma')s[Y(e,x,\gamma_n^\circ),x,\lambda_n^\circ]$ can be deduced as follows

$$\begin{aligned}
&\mathcal{E}(\partial^2/\partial\sigma_i\partial\sigma_j)s[Y(e,x,\gamma_n^\circ),x,\lambda_n^\circ] \\
&\qquad = -\tfrac{1}{2}\mathrm{tr}\ (\Sigma_n^\circ)^{-1}\Sigma(\xi_j)(\Sigma_n^\circ)^{-1}\Sigma(\xi_i)[I - 2(\Sigma_n^\circ)^{-1}(\Sigma_n^\circ)] \\
&\qquad = \tfrac{1}{2}\ \mathrm{vec}'[\Sigma(\xi_j)(\Sigma_n^\circ)^{-1}][I \otimes (\Sigma_n^\circ)^{-1}]\mathrm{vec}\ \Sigma(\xi_j) \\
&\qquad = \tfrac{1}{2}\ \xi_j'\ K'(\Sigma_n^\circ \otimes \Sigma_n^\circ)^{-1}K\ \xi_i\ .
\end{aligned}$$

Since $\mathcal{E}(\partial^2/\partial\sigma_i\partial\sigma_j)s[Y(e,x,\gamma_n^\circ),x,\lambda_n^\circ] = 0$ we have

$$\mathcal{J}_n^\circ = \begin{pmatrix} \Omega_n^\circ & 0 \\ 0 & \tfrac{1}{2}K'(\Sigma_n^\circ \otimes \Sigma_n^\circ)^{-1}K \end{pmatrix} .$$

Normality plays no role in the form of $\mathcal{J}_n^\circ$.

In summary we have

NOTATION 7a (in general)

$$\mathcal{I}_n^\circ = \begin{pmatrix} \Omega_n^\circ & (\mathfrak{J}_n^\circ)'[(\Sigma_n^\circ)^{-\frac{1}{2}}]'\mathcal{E}[e\ \mathrm{vec}'(ee')](\Sigma_n^\circ \otimes \Sigma_n^\circ)^{-\frac{1}{2}}K \\ \mathrm{sym} & \frac{1}{4}K'[(\Sigma_n^\circ \otimes \Sigma_n^\circ)^{-\frac{1}{2}}]'\mathrm{Var}[\mathrm{vec}(ee')](\Sigma_n^\circ \otimes \Sigma_n^\circ)^{-\frac{1}{2}}K \end{pmatrix}$$

$$\mathcal{J}_n^\circ = \begin{pmatrix} \Omega_n^\circ & 0 \\ 0 & \frac{1}{2}K'(\Sigma_n^\circ \otimes \Sigma_n^\circ)^{-1}K \end{pmatrix}$$

$$\mathcal{U}_n^\circ = 0$$

NOTATION 7b (under normality)

$$\mathcal{I}_n^\circ = \begin{pmatrix} \Omega_n^\circ & 0 \\ 0 & \frac{1}{2}K'(\Sigma_n^\circ \otimes \Sigma_n^\circ)^{-1}K \end{pmatrix}$$

$$\mathcal{J}_n^\circ = \mathcal{I}_n^\circ$$

$$\mathcal{U}_n^\circ = 0.$$

The expressions for $\mathcal{I}^*$, $\mathcal{J}^*$, and $\mathcal{U}^*$ have the same form as above with $(\Omega, \mathfrak{J}, \Sigma^*)$ replacing $(\Omega_n^\circ, \mathfrak{J}_n^\circ, \Sigma_n^\circ)$ throughout. □

Let $\hat{\lambda}_n = (\hat{\theta}_\infty, \hat{\Sigma}_\infty)$ denote the minimum of $s_n(\theta,\Sigma)$ and let $\tilde{\lambda}_n = (\tilde{\theta}_\infty, \tilde{\Sigma}_\infty)$ denote the minimum of $s_n(\theta,\Sigma)$ subject to $h(\lambda) = 0$. Define:

NOTATION 8.

$$\hat{\Omega} = (1/n)\Sigma_{t=1}^{n}[(\partial/\partial\theta')f(x_t,\hat{\theta}_\infty)]'\hat{\Sigma}_\infty^{-1}[(\partial/\partial\theta')f(x_t,\hat{\theta}_\infty)]$$

$$\hat{S}_t = \begin{pmatrix} [(\partial/\partial\theta')f(x_t,\hat{\theta}_\infty)]'\Sigma_\infty^{-1}\hat{u}_t \\ \tfrac{1}{2}K'(\hat{\Sigma}_\infty \otimes \hat{\Sigma}_\infty)^{-1}\operatorname{vec}[\hat{u}_t\hat{u}_t' - \hat{\Sigma}_\infty^{-1}] \end{pmatrix}$$

$$\hat{u}_t = y_t - f(x_t,\hat{\theta}_\infty)$$

Expressions for $\tilde{\Omega}$, $\tilde{S}_t$, and $\tilde{u}_t$ are the same with $(\tilde{\theta}_\infty,\tilde{\Sigma}_\infty)$ replacing $(\hat{\theta}_\infty,\hat{\Sigma}_\infty)$ throughout. []

We propose the following as estimators of $\mathcal{I}^*$ and $\mathcal{J}^*$.

NOTATION 9a (in general)

$$\hat{\mathcal{I}} = (1/n)\Sigma_{t=1}^{n}\hat{S}_t\hat{S}_t'$$

$$\hat{\mathcal{J}} = \begin{pmatrix} \hat{\Omega} & 0 \\ 0 & \tfrac{1}{2}K'(\hat{\Sigma}_\infty \otimes \hat{\Sigma}_\infty)^{-1}K \end{pmatrix}$$

NOTATION 9b (under normality)

$$\hat{\mathcal{I}} = \hat{\mathcal{J}} = \begin{pmatrix} \hat{\Omega} & 0 \\ 0 & \tfrac{1}{2}K'(\hat{\Sigma}_\infty \otimes \hat{\Sigma}_\infty)^{-1}K \end{pmatrix}$$

The expressions for $\tilde{\mathcal{I}}$ and $\tilde{\mathcal{J}}$ have the same form with $(\tilde{\Omega},\tilde{\Sigma},\tilde{S}_t)$ replacing $(\hat{\Omega},\hat{\Sigma},\hat{S}_t)$ throughout. []

A test of the marginal hypothesis

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

where $h(\theta)$ maps $\mathbb{R}^p$ into $\mathbb{R}^q$ is most often of interest in applications. As mentioned earlier, maximum likelihood estimators are least squares estimators so that, as regards the Wald and the Lagrange multiplier tests, the theory and methods set forth in Section 4 can be applied directly with the maximum likelihood estimators

$$\hat{\theta}_\infty, \hat{\Sigma}_\infty, \tilde{\theta}_\infty, \tilde{\Sigma}_\infty$$

replacing, respectively, the estimators

$$\hat{\theta},\ \hat{\Sigma},\ \tilde{\tilde{\theta}},\ \tilde{\Sigma}$$

in the formulas for the Wald and Lagrange multiplier test statistics. The likelihood ratio test needs modification due to the following considerations.

Direct application of Theorem 15 of Chapter 3 would give

$$\begin{aligned} L1 &= 2n[s_n(\tilde{\theta}_\infty, \tilde{\Sigma}_\infty) - s_n(\hat{\theta}_\infty, \hat{\Sigma}_\infty)] \\ &= n(\ell n \det \tilde{\Sigma}_\infty - \ell n \det \hat{\Sigma}_\infty) \end{aligned}$$

as the likelihood ratio test statistic whereas application of the results in Section 4 would give

$$\begin{aligned} L2 &= \frac{[S(\tilde{\theta}, \hat{\Sigma}_\infty) - S(\hat{\theta}_\infty, \hat{\Sigma}_\infty)]/q}{S(\hat{\theta}_\infty, \hat{\Sigma}_\infty)/(nM-p)} \\ &= \frac{[S(\tilde{\theta}, \hat{\Sigma}_\infty) - nM]/q}{nM/(nM-p)} \quad . \end{aligned}$$

where $\tilde{\theta}$ minimize $S(\theta, \hat{\Sigma}_\infty)$ subject to $h(\theta) = 0$. These two formulas can be reconciled using the equation

$$d\ \ell n \det \Sigma = \operatorname{tr}(\Sigma^{-1} d\Sigma)$$

or

$$\ell n \det(\Sigma + \Delta) - \ell n \det \Sigma = \operatorname{tr}(\Sigma^{-1}\Delta) + o(\Delta)$$

derived in Problem 4.

To within a differential approximation

$$S(\tilde{\theta},\hat{\Sigma}_\infty) - S(\hat{\theta}_\infty,\hat{\Sigma}_\infty)$$

$$= n\,\mathrm{tr}[\hat{\Sigma}_\infty^{-1}\,\hat{\Sigma}(\tilde{\theta})] - n\,\mathrm{tr}(\hat{\Sigma}_\infty^{-1}\hat{\Sigma}_\infty)$$

$$= n\,\mathrm{tr}\,\hat{\Sigma}_\infty^{-1}[\hat{\Sigma}(\tilde{\theta}) - \hat{\Sigma}_\infty]$$

$$= n\,\mathrm{tr}\,\hat{\Sigma}_\infty^{-1}\,(\tilde{\Sigma}_\infty - \hat{\Sigma}_\infty) + n\,\mathrm{tr}\,\hat{\Sigma}_\infty^{-1}[\hat{\Sigma}(\tilde{\theta}) - \tilde{\Sigma}_\infty]$$

$$\doteq L1 + n\,\mathrm{tr}\,\hat{\Sigma}_\infty^{-1}[\hat{\Sigma}(\tilde{\theta}) - \tilde{\Sigma}_\infty]$$

$$= L1 + [S(\tilde{\theta},\hat{\Sigma}_\infty) - S(\tilde{\theta}_\infty,\hat{\Sigma}_\infty)].$$

Thus one can expect that there will be a negligible difference between an inference based on either L1 or L2 in most applications. Our recommendation is to use $L1 = n(\ell n\det\tilde{\Sigma}_\infty - \ell n\det\hat{\Sigma}_\infty)$ to avoid the confusion that would result from the use of something other than the classical likelihood ratio test in connection with maximum likelihood estimators. But we do recommend the use of degrees of freedom corrections to improve the accuracy of probability statements.

To summarize this discussion, the likelihood ratio test rejects the hypothesis

$$H\colon\ h(\theta^\circ) = 0$$

where $h(\theta)$ maps $\mathbb{R}^p$ into $\mathbb{R}^q$ when the statistic

$$L = n\,(\ell n\det\tilde{\Sigma}_\infty - \ell n\det\hat{\Sigma}_\infty)$$

exceeds qF_α where F_α denotes the upper $\alpha \times 100\%$ critical point of the F-distribution with q numerator degrees of freedom and nM-p denominator degrees of freedom; $F_\alpha = F^{-1}(1-\alpha; q, nM-p)$.

We illustrate.

EXAMPLE 1. (continued) Consider retesting the hypothesis of homogeneity, expressed as the functional dependency

$$H\colon \theta^{\circ} = g(\rho^{\circ}) \text{ for some } \rho^{\circ} \text{ against } A\colon \theta^{\circ} \neq g(\rho) \text{ for any } \rho$$

with

$$g(\rho) = \begin{pmatrix} \rho_1 \\ -\rho_2-\rho_3 \\ \rho_2 \\ \rho_3 \\ \rho_4 \\ -\rho_5-\rho_2 \\ \rho_5 \\ -\rho_5-\rho_3 \end{pmatrix},$$

in the model with response function

$$f(x,\theta) = \begin{pmatrix} \ell n \dfrac{\theta_1 + \theta_2 x_1 + \theta_3 x_2 + \theta_4 x_3}{-1 + \theta_4 x_1 + \theta_7 x_2 + \theta_8 x_3} \\[2ex] \ell n \dfrac{\theta_5 + \theta_3 x_1 + \theta_6 x_2 + \theta_7 x_3}{-1 + \theta_4 x_1 + \theta_7 x_2 + \theta_8 x_3} \end{pmatrix}$$

using the likelihood ratio test; θ has length $p = 8$ and ρ has length $r = 5$ whence $q = p - r = 3$. The model is bivariate so $M = 2$ and there are $n = 224$ observations.

In Figure 10a the maximum likelihood estimators are computed by iterating the least squares estimator to convergence obtaining

$$\hat{\theta}_{\infty} = \begin{pmatrix} -2.92345 \\ -1.28826 \\ 0.81849 \\ 0.36121 \\ -1.53759 \\ -1.04926 \\ 0.02987 \\ -0.46741 \end{pmatrix} \qquad \text{(from Figure 10a)}$$

$$\hat{\Sigma}_{\infty} = \begin{pmatrix} 0.165141 & 0.92505 \\ 0.092505 & 0.08989 \end{pmatrix} \qquad \text{(from Figure 10a).}$$

Compare these values with those shown in Figures 3b and 3c; the difference is slight.

Figure 10a. Example 1 Fitted by Maximum Likelihood, Across Equation Constraints Imposed.

SAS Statements:

```
PROC MODEL OUT=MODEL01;
ENDOGENOUS Y1 Y2;
EXOGENOUS  X1 X2 X3;
PARMS T1 -2.98  T2 -1.16  T3 0.787  T4 0.353  T5 -1.51  T6 -1.00
      T7 0.054  T8 -0.474;
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);
PROC SYSNLIN DATA=EXAMPLE1  MODEL=MODEL01 ITSUR NESTIT METHOD=GAUSS OUTS=SHAT;
```

Output:

NONLINEAR ITSUR PARAMETER ESTIMATES

PARAMETER	ESTIMATE	APPROX. STD ERROR	'T' RATIO	APPROX. PROB>|T|
T1	-2.92345	0.2781988	-10.51	0.0001
T2	-1.28826	0.226682	-5.68	0.0001
T3	0.8184883	0.08079815	10.13	0.0001
T4	0.3612072	0.03033416	11.91	0.0001
T5	-1.53759	0.09204265	-16.71	0.0001
T6	-1.04926	0.08368577	-12.54	0.0001
T7	0.02986769	0.03617161	0.83	0.4099
T8	-0.467411	0.01927753	-24.25	0.0001

SYSTEM STATISTICS: SSE = 447.9999 MSE = 2 OBS= 224

COVARIANCE OF RESIDUALS

	Y1	Y2
Y1	0.165141	0.0925046
Y2	0.0925046	0.0898862

In Figure 10b the estimator $\hat{\rho}_\infty$ minimizing $\ell n \det \Sigma[g(\rho)]$ is obtained by iterated least squares; put $\tilde{\theta}_\infty = g(\hat{\rho}_\infty)$ to obtain

$$\tilde{\theta}_\infty = g(\hat{\rho}_\infty) = \begin{pmatrix} -2.7303 \\ -1.2315 \\ 0.8582 \\ 0.3733 \\ -1.5935 \\ -0.9167 \\ 0.0585 \\ -0.4319 \end{pmatrix} \qquad \text{(from Figure 10b)}$$

$$\tilde{\Sigma}_\infty = \begin{pmatrix} 0.179194 & 0.095365 \\ 0.095365 & 0.090199 \end{pmatrix} \qquad \text{(from Figure 10b)}$$

Compare these values with those shown in Figure 6; again, the difference is slight.

In Figure 10c, the likelihood ratio test statistic is computed as

$$L = n(\ell n \det \tilde{\Sigma}_\infty - \ell n \det \hat{\Sigma}_\infty)$$

$$= 26.2573 \qquad \text{(from Figure 10c).}$$

$F_\alpha = F^{-1}(.95;3,440) = 2.61$ so that

$$q\, F_\alpha = (3)(2.61) = 7.83.$$

One rejects

$$H\colon\ \theta^\circ = g(\rho^\circ) \text{ for some } \rho^\circ$$

at the 5% level. With this many denominator degrees of freedom, the difference between $q\, F_\alpha$ and the three degrees of freedom chi-square critical value of 7.81 is negligible. In smaller sized samples this will not be the case.

It is of interest to compare

$$n(\ell n \det \tilde{\Sigma}_\infty - \ell n \det \hat{\Sigma}_\infty) = 26.2573 \text{ (from Figure 10c)}$$

with

$$S(\tilde{\theta},\hat{\Sigma}) - S(\hat{\theta},\hat{\Sigma})$$

$$= 474.6822 - 446.8570 \quad \text{(from Figures 6 and 3c)}$$

$$= 27.8252\ .$$

Figure 10b. Example 1 Fitted by Maximum Likelihood, Across Equation Constraints Imposed, Homogeneity Imposed.

SAS Statements:

```
PROC MODEL OUT=MODEL02;
ENDOGENOUS Y1 Y2;
EXOGENOUS  X1 X2 X3;
PARMS R1 -2.72  R2 0.858  R3 0.374  R4 -1.59  R5 0.057;
T1=R1; T2=-R2-R3; T3=R2; T4=R3; T5=R4; T6=-R5-R2; T7=R5; T8=-R5-R3;
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);
PROC SYSNLIN DATA=EXAMPLE1  MODEL=MODEL02 ITSUR NESTIT METHOD=GAUSS OUTS=STILDE;
```

Output:

NONLINEAR ITSUR PARAMETER ESTIMATES

PARAMETER	ESTIMATE	APPROX. STD ERROR	'T' RATIO	APPROX. PROB>|T|
R1	-2.7303	0.1800188	-15.17	0.0001
R2	0.8581672	0.06691972	12.82	0.0001
R3	0.3733482	0.02736511	13.64	0.0001
R4	-1.59345	0.07560367	-21.08	0.0001
R5	0.05854239	0.0339787	1.72	0.0863

SYSTEM STATISTICS: SSE = 448 MSE = 2 OBS= 224

COVARIANCE OF RESIDUALS

	Y1	Y2
Y1	0.179194	0.0953651
Y2	0.0953651	0.0901989

Figure 10c. Illustration of Likelihood Ratio Test Computations with Example 1.

SAS Statements:

```
PROC MATRIX;
FETCH SHAT DATA=SHAT(KEEP=Y1 Y2);
FETCH STILDE DATA=STILDE(KEEP=Y1 Y2);
N=224;  L=N#(LOG(DET(STILDE))-LOG(DET(SHAT)));  PRINT L;
```

Output:

```
                         SAS                                11


           L                  COL1

           ROW1            26.2573
```

The differential approximation $d\,\ell n \det \Sigma = \operatorname{tr} \Sigma^{-1} d\Sigma$ seems to be reasonably accurate in this instance. □

A marginal hypothesis of the form

$$H\colon h(\sigma^\circ) = 0 \text{ against } A\colon h(\sigma^\circ) \neq 0$$

is sometimes of interest in applications. We shall proceed under the assumption that the computation of $(\hat{\theta}_\infty, \hat{\sigma}_\infty)$ is fairly straightforward but that the minimization of $s_n(\theta,\sigma)$ subject to $h(\sigma) = 0$ is inordinately burdensome as is quite often the case. This assumption compels the use of the Wald test statistic. We shall also assume that the errors are normally distributed.

Under normality, the Wald test statistic for the hypothesis

$$H\colon h(\sigma^\circ) = 0 \text{ against } A\colon h(\sigma^\circ) \neq 0$$

where $h(\sigma)$ maps $\mathbb{R}^{M(M+1)/2}$ into $\mathbb{R}^q$ has the form

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

where

$$\hat{h} = h(\hat{\sigma}_\infty)\ ,$$

$$\hat{H} = (\partial/\partial\sigma')\, h(\hat{\sigma}_\infty)\ ,$$

$$\hat{V} = [\tfrac{1}{2}K'(\hat{\Sigma}_\infty \otimes \hat{\Sigma}_\infty)^{-1}K]^{-1}\ .$$

The test rejects when W exceeds the upper $\alpha \times 100\%$ critical point of a chi-square random variable with q degrees of freedom.

In performing the computations, explicit construction of the matrix K can be avoided as follows. Consider w defined by

$$\operatorname{vec} uu' = \Sigma(w) = Kw$$

where u is an M-vector. Subscripts are related as follows

$$\text{vec } uu' = \begin{bmatrix} u_1u_1 \\ u_1u_2 \\ \vdots \\ u_\alpha u_\beta \\ \vdots \\ u_M u_{M-1} \\ u_M u_M \end{bmatrix} \xleftarrow{} i = \beta(\beta-1)/2 + \alpha \xrightarrow{} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_i \\ \vdots \\ w_{M(M+1)/2} \end{bmatrix} .$$

If $u \sim N_M(0,\Sigma)$ then for

$$i = \beta(\beta-1)/2 + \alpha$$

$$j = \beta'(\beta'-1)/2 + \alpha'$$

we have (Anderson, 1958, p. 161) that

$$\mathcal{C}(w_i,w_j) = \mathcal{E}(u_\alpha u_\beta - \sigma_{\alpha\beta})(u_{\alpha'}u_{\beta'} - \sigma_{\alpha'\beta'})$$

$$= \sigma_{\alpha\alpha'}\,\sigma_{\beta\beta'} + \sigma_{\alpha\beta'}\,\sigma_{\beta\alpha'} .$$

Thus, the variance-covariance matrix $\mathcal{C}(w,w')$ of the random variable w can be computed easily. Now consider the asymptotics for the model $y_t = u_t = \Sigma^{\frac{1}{2}} e_t$ with e_t independent $N(0,I)$. The previous asymptotic results imply

$$\sqrt{n}(\hat{\sigma}_\infty - \sigma) \xrightarrow{\mathcal{L}} N\{0,[\tfrac{1}{2}K(\Sigma \otimes \Sigma)^{-1}K\;]^{-1}\}$$

but in this case $\hat{\sigma}_\infty = (1/n)\Sigma_{t=1}^{n} w_t$ and the Central Limit Theorem implies that

$$\sqrt{n}(\hat{\sigma}_\infty - \sigma) \xrightarrow{\mathcal{L}} N[0,\mathcal{C}(w,w')] .$$

We conclude that

$$V = [\tfrac{1}{2}K(\Sigma \otimes \Sigma)^{-1}K']^{-1} = \mathcal{C}(w,w')$$

and have the following algorithm for computing the elements v_{ij} of V.

```
DO for β = 1 to M;
DO for α = 1 to β;
i = β(β-1)/2 + α;
   DO for β' = 1 to M;
   DO for α' = 1 to β';
   j = β'(β'-1)/2 + α';
      v_ij = σ_αα' σ_ββ' + σ_αβ' σ_βα';
   END;
   END;
END;
END;
```

Table 5. Yields of 3 Varieties of Alfalfa (Tons Per Acre) in 1944 Following 4 Dates of Final Cutting in 1943.

			Date			
t	Variety	Block	A	B	C	D
1	Ladak	1	2.17	1.58	2.29	2.23
2	Cossac	1	2.33	1.38	1.86	2.27
3	Ranger	1	1.75	1.52	1.55	1.56
4	Ladak	2	1.88	1.26	1.60	2.01
5	Cossac	2	2.01	1.30	1.70	1.81
6	Ranger	2	1.95	1.47	1.61	1.72
7	Ladak	3	1.62	1.22	1.67	1.82
8	Cossac	3	1.70	1.85	1.81	2.01
9	Ranger	3	2.13	1.80	1.82	1.99
10	Ladak	4	2.34	1.59	1.91	2.10
11	Cossac	4	1.78	1.09	1.54	1.40
12	Ranger	4	1.78	1.37	1.56	1.55
13	Ladak	5	1.58	1.25	1.39	1.66
14	Cossac	5	1.42	1.13	1.67	1.31
15	Ranger	5	1.31	1.01	1.23	1.51
16	Ladak	6	1.66	0.94	1.12	1.10
17	Cossac	6	1.35	1.06	0.88	1.06
18	Ranger	6	1.30	1.31	1.13	1.33

Source: Snedecor and Cochran(1980)

We illustrate with an example.

EXAMPLE 2. (Split Plot Design) The split plot experimental design can be viewed as a two-way design with multivariate observations in each cell which is written as

$$y_{ij} = u + \rho_i + \tau_j + e_{ij}$$

where y_{ij}, u, etc. are M-vectors and

$$i = 1, 2, \ldots, I = \# \text{ blocks}$$

$$j = 1, 2, \ldots, J = \# \text{ treatments}$$

$$\mathcal{C}(e_{ij}e'_{ij}) = \Sigma \ .$$

In the corresponding univariate split plot analysis, the data are assumed to follow the model

$$y_{kij} = m + r_i + t_j + \eta_{ij} + s_k + (rs)_{ki} + (ts)_{kj} + \varepsilon_{kij}$$

where k = 1, 2, ..., M, Roman letters denote parameters, and Greek letters denote random variables, $\text{Var}(\eta_{ij}) = \sigma^2_\eta$, $\text{Var}(\varepsilon_{kij}) = \sigma^2_\varepsilon$, and random variables with different subscripts are assumed independent. It is not difficult to show (Problem 5) that the only difference between the two models is that the univariate analysis imposes the restriction

$$\Sigma = \sigma^2_\varepsilon I + \sigma^2_\eta J$$

on the variance-covariance matrix of the multivariate analysis; I is the identity matrix of order M and J is an M by M matrix whose entries are all ones. Such an assumption is somewhat suspect when the observations

$$y'_{ij} = (y_{1ij}, y_{2ij}, \ldots, y_{Mij})$$

represent successive observations on the same plot at different points in time. An instance is the data shown in Table 5.

For these data, $M = 4$, $n = 18$, and the hypothesis to be tested is H: $h(\sigma^{\circ}) = 0$ where

$$h(\sigma) = \underbrace{\begin{pmatrix} 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \end{pmatrix}}_{H} \underbrace{\begin{pmatrix} \sigma_{11} \\ \sigma_{12} \\ \sigma_{22} \\ \sigma_{13} \\ \sigma_{23} \\ \sigma_{33} \\ \sigma_{14} \\ \sigma_{24} \\ \sigma_{34} \\ \sigma_{44} \end{pmatrix}}_{\sigma}$$

Figure 11. Maximum Likelihood Estimation of the Variance-Covariance Matrix of Example 2.

SAS Statements:

```
PROC ANOVA DATA=EXAMPLE2;
CLASSES VARIETY BLOCK;
MODEL A B C D = VARIETY BLOCK;
MANOVA / PRINTE;
```

Output:

SAS 6

ANALYSIS OF VARIANCE PROCEDURE

E = ERROR SS&CP MATRIX

DF=10	A	B	C	D
A	0.56965556	0.23726111	0.25468889	0.36578889
B	0.23726111	0.46912222	0.26341111	0.31137778
C	0.25468889	0.26341111	0.42495556	0.25678889
D	0.36578889	0.31137778	0.25678889	0.60702222

The maximum likelihood estimate of Σ is computed in Figure 11 as

$$\hat{\Sigma}_\infty = \begin{pmatrix} 0.0316475 & 0.0131812 & 0.0141494 & 0.0203216 \\ 0.0131812 & 0.0260623 & 0.0146340 & 0.0172988 \\ 0.0141494 & 0.0146340 & 0.0236086 & 0.0142660 \\ 0.0203216 & 0.0172988 & 0.0142660 & 0.0337235 \end{pmatrix} = \Sigma(\hat{\sigma}_\infty)$$

whence

$$h(\hat{\sigma}_\infty) = \begin{pmatrix} 0.00558519 \\ 0.00803889 \\ -0.00207593 \\ -0.00096821 \\ -0.00145278 \\ -0.00714043 \\ -0.00411759 \\ -0.00108488 \end{pmatrix} .$$

Figure 12. Wald Test of a Restriction on the Variance-Covariance Matrix of Example 2.

SAS Statements:

```
PROC MATRIX;
SSCP =  0.56965556  0.23726111  0.25468889  0.36578889/
        0.23726111  0.46912222  0.26341111  0.31137778/
        0.25468889  0.26341111  0.42495556  0.25678889/
        0.36578889  0.31137778  0.25678889  0.60702222;
S    = (0.56965556  0.23726111  0.46912222  0.25468889  0.26341111
        0.42495556  0.36578889  0.31137778  0.25678889  0.60702222)';
HH   =  1  0  -1   0  0  0  0  0  0  0/
        1  0   0   0  0 -1  0  0  0  0/
        1  0   0   0  0  0  0  0  0 -1/
        0  1   0  -1  0  0  0  0  0  0/
        0  1   0   0 -1  0  0  0  0  0/
        0  1   0   0  0  0 -1  0  0  0/
        0  1   0   0  0  0  0 -1  0  0/
        0  1   0   0  0  0  0  0 -1  0;
N=18;  SIGMA=SSCP#/N;  H=HH*S#/N;
M=4; V=(0*(1:M*(M+1)#/2))'*(0*(1:M*(M+1)#/2));
DO  B = 1 TO M;  DO  A = 1 TO  B;  I =  B*( B-1)#/2+ A;
DO BB = 1 TO M;  DO AA = 1 TO BB;  J = BB*(BB-1)#/2+AA;
V(I,J)=SIGMA(A,AA)*SIGMA(B,BB)+SIGMA(A,BB)*SIGMA(B,AA);
END;  END;  END;  END;
WALD=N*(H'*INV(HH*V*HH')*H);  PRINT WALD;
```

Output:

```
                                SAS                                1


                      WALD          COL1

                      ROW1        2.26972
```

Figure 13 illustrates the algorithm for computing $\hat{V}$ discussed above and we obtain

$$W = 2.26972 \qquad \text{(from Figure 12).}$$

Entering a table of the chi-square distribution at 8 degrees of freedom one finds that

$$p = P(W > 2.26972) \doteq 0.97 \; .$$

A univariate analysis of the data seems reasonable.

This happens to be an instance where it is easy to compute the maximum likelihood estimate subject to $h(\sigma) = 0$. From Figure 13 we obtain

$$\tilde{\Sigma}_\infty = \begin{pmatrix} 0.0287605 & 0.0156418 & 0.0156418 & 0.0156418 \\ 0.0156418 & 0.0287605 & 0.0156418 & 0.0156418 \\ 0.0156418 & 0.0156418 & 0.0287605 & 0.0156418 \\ 0.0156418 & 0.0156418 & 0.0156418 & 0.0287605 \end{pmatrix} .$$

For a linear model of this form we have (Problem 6)

$$\begin{aligned} L &= 2n[s_n(\tilde{\theta}_\infty, \tilde{\Sigma}_\infty) - s_n(\hat{\theta}_\infty, \hat{\Sigma}_\infty)] \\ &= n[\ell n \det \tilde{\Sigma}_\infty + \operatorname{tr} \tilde{\Sigma}_\infty^{-1} \hat{\Sigma}_\infty - \ell n \det \hat{\Sigma}_\infty - M] \\ &= 2.61168 \quad \text{(from Figure 14)} \end{aligned}$$

Figure 13. Maximum Likelihood Estimation of the Variance-Covariance Matrix of Example 2 under the ANOVA Restriction.

SAS Statements:

```
PROC VARCOMP DATA=EXAMPLE2 METHOD=ML;
CLASSES VARIETY DATE BLOCK;
MODEL
YIELD = BLOCK VARIETY DATE DATE*BLOCK DATE*VARIETY BLOCK*VARIETY /
FIXED = 5;
```

Output:

```
                              SAS                                    2

     MAXIMUM LIKELIHOOD VARIANCE COMPONENT ESTIMATION PROCEDURE

DEPENDENT VARIABLE: YIELD

     ITERATION      OBJECTIVE  VAR(VARIETY*BLOCK)          VAR(ERROR)

         0      -280.48173508          0.01564182          0.01311867
         1      -280.48173508          0.01564182          0.01311867

     CONVERGENCE CRITERION MET
```

Figure 14. Likelihood Ratio Test of a Restriction on the Variance-Covariance Matrix of Example 2.

SAS Statements:

```
PROC MATRIX;
SSCP  = 0.56965556  0.23726111  0.25468889  0.36578889/
        0.23726111  0.46912222  0.26341111  0.31137778/
        0.25468889  0.26341111  0.42495556  0.25678889/
        0.36578889  0.31137778  0.25678889  0.60702222;
N=18;  M=4;  SHAT=SSCP#/N;  STILDE=0.01311867#I(M)+J(M,M,0.01564182);
L=N#(LOG(DET(STILDE))+TRACE(INV(STILDE)*SHAT)-LOG(DET(SHAT))-M);  PRINT L;
```

Output:

```
                                SAS                                    1


                  L              COL1

                  ROW1        2.61168
```

which agrees well with the Wald test statistic. []

A test of a joint hypothesis

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

is not encountered very often in applications. In the event that it is, application of Theorems 11, 14, or 15 of Section 5, Chapter 3 is reasonably straightforward.

PROBLEMS

1. Show that $\sqrt{n}\,(\hat{\Sigma}_I - \Sigma_n^\circ)$ is bounded in probability. Hint, see Problem 1, Section 3.

2. Show that $(\partial/\partial\sigma_i)\Sigma(\sigma) = \Sigma(\xi_i)$. In Problem 4 the expressions $(\partial/\partial\sigma_i)\,\Sigma^{-1}(\sigma) = -\Sigma^{-1}(\sigma)[(\partial/\partial\sigma_i)\Sigma(\sigma)]\Sigma^{-1}(\sigma)$ and $(\partial/\partial\sigma_i)\,\ell n \det \Sigma(\sigma) = \mathrm{tr}[\Sigma^{-1}(\sigma)(\partial/\partial\sigma_i)\Sigma(\sigma$ are derived. Use them to derive the first and second partial derivatives of $s[Y(e,x,\gamma^\circ),x,\lambda]$ given in the text.

3. Denote the j^{th} column of a matrix A by $A_{(j)}$ and a typical element by a_{ij}. Show that

$$
\begin{aligned}
(ABC)_{(j)} &= (AB)C_{(j)} \\
&= \Sigma_i (c_{ij}\,A)(B_{(i)}) \\
&= [(C_{(j)})' \otimes A]\ \mathrm{vec}\ B
\end{aligned}
$$

then stack the columns $(ABC)_{(j)}$ to obtain

$$\mathrm{vec}(ABC) = (C' \otimes A)\mathrm{vec}\ B.$$

Show that

$$\mathrm{tr}(AB) = \Sigma_i \Sigma_k a_{ik} b_{ki} = \mathrm{vec}'(A')\mathrm{vec}\ B$$

whence

$$\mathrm{tr}(ABC) = \mathrm{vec}'(A')\mathrm{vec}(BCI) = \mathrm{vec}'(A')(I \otimes B)\mathrm{vec}\ C.$$

4. Show that $(\partial/\partial\sigma_i)\Sigma(\sigma) = \Sigma(\xi_i)$. Use $I = \Sigma^{-1}(\sigma)\Sigma(\sigma)$ to obtain $0 = [(\partial/\partial\sigma_i)\Sigma^{-1}(\sigma)]\Sigma(\sigma) + \Sigma^{-1}(\sigma)[(\partial/\partial\sigma_i)\Sigma(\sigma)]$ whence $(\partial/\partial\sigma_i)\Sigma^{-1}(\sigma) = -\Sigma^{-1}(\sigma)[(\partial/\partial\sigma_i)\Sigma(\sigma)]\Sigma^{-1}(\sigma)$. Let a square matrix A have elements a_{ij}, let c_{ij} denote the cofactors of A and let a^{ij} denote the elements of A^{-1}. From $\det A = \Sigma_k\, a_{ik}c_{ik}$ show that $(\partial/\partial a_{ij}) \det A = c_{ij} = a^{ji} \det A$. This implies that

$$(\partial/\partial\, \mathrm{vec}'A)\ \det A = \det A\ \mathrm{vec}'(A^{-1})' .$$

Use this fact and the previous problem to show

$$(\partial/\partial\sigma_i)\ \det \Sigma(\sigma) = \det \Sigma(\sigma)\ \mathrm{vec}'[\Sigma^{-1}(\sigma)]\ \mathrm{vec}[(\partial/\partial\sigma_i)\Sigma(\sigma)]$$

$$= \det \Sigma(\sigma)\ \mathrm{tr}[\Sigma^{-1}(\sigma)(\partial/\partial\sigma_i)\Sigma(\sigma)].$$

Show that $(\partial/\partial\sigma_i)\ \ell n \det \Sigma(\sigma) = \mathrm{tr}[\Sigma^{-1}(\sigma)(\partial/\partial\sigma_i)\Sigma(\sigma)]$.

5. Referring to Example 2, a two-way multivariate design has fixed part

$$\mathcal{E}y_{ij} = \mu + \rho_i + \tau_j$$

whereas the split-plot ANOVA has fixed part

$$\mathcal{E}y_{ijk} = m + r_i + t_j + s_k + (rs)_{ki} + (ts)_{kj} .$$

Use the following correspondences to show that the fixed part of the designs is the same

$$\mu' = (\mu_1, \ldots, \mu_k, \ldots, \mu_M) = (m, \ldots, m, \ldots, m) ,$$

$$\rho_i' = (\rho_{1i},\ldots,\rho_{ki},\ldots,\rho_{Mi}) = (r_i + (rs)_{1i},\ldots,r_i + (rs)_{ki},\ldots,r_i + (rs)_{Mi}) ,$$

$$\tau_j' = (\tau_{1j},\ldots,\tau_{kj},\ldots,\tau_{Mj}) = (t_j + (ts)_{1j},\ldots,t_j + (ts)_{kj},\ldots,t_j + (ts)_{Mj}) .$$

Show that under the ANOVA assumption $\Sigma = \sigma_\varepsilon^2\, I + \sigma_\eta^2\, J$.

6. Suppose that one has the multivariate linear model

$$y_t' = x_t'B + e_t' \qquad t = 1, 2, \ldots, n .$$

where B is k by p and y_t' is 1 by M. Write

$$Y = \begin{pmatrix} y_1' \\ y_2' \\ \vdots \\ y_n' \end{pmatrix} , \qquad X = \begin{pmatrix} x_1' \\ x_2' \\ \vdots, \\ x_n' \end{pmatrix}$$

and show that

$$2\, s_n(B,\Sigma) = \ell n \det \Sigma + \operatorname{tr} \Sigma^{-1}[Y - P_X Y]'[Y - P_X Y]/n$$
$$+ \operatorname{tr} \Sigma^{-1}[P_X Y - XB]'[P_X Y - XB]/n$$

where $P_X = X(X'X)^{-1}X'$. One observes from this equation that $\hat{B}$ will be computed as $\hat{B} = (X'X)^{-1}X'Y$ no matter what value is assigned to Σ. Thus, if $(\tilde{B}_\infty, \tilde{\Sigma}_\infty)$ minimizes $s_n(B,\Sigma)$ subject to $\Sigma = \Sigma(\sigma)$, $h(\sigma) = 0$ and $(\hat{B}_\infty, \hat{\Sigma}_\infty)$ is the unconstrained minimizer then

$$2\, s_n(\tilde{B}_\infty, \tilde{\Sigma}_\infty) = \ell n \det \tilde{\Sigma}_\infty + \operatorname{tr} \tilde{\Sigma}_\infty^{-1} \hat{\Sigma}_\infty + 0 \; .$$

7. AN ILLUSTRATION OF THE BIAS IN INFERENCE CAUSED BY MISSPECIFICATION

The asymptotic theory in Chapter 3 was developed in sufficient generality to permit the analysis of inference procedures under conditions where the data generating model and the model fitted to the data are not the same. The following example is an instance where a second order polynomial approximation can lead to considerable error. The underlying ideas are similar to those of Example 1.

EXAMPLE 3. (Power curve of a Translog test of additivity) The theory of demand states that of the bundles of goods and services that the consumer can afford he will choose that bundle which pleases him the most. Mathematically this proposition is stated as follows: Let there be N different goods and services in a bundle, let $q = (q_1, q_2, \ldots, q_N)'$ be an N-vector giving the quantities of each in a bundle, let $p = (p_1, p_2, \ldots, p_N)'$ be the N-vector of corresponding prices, let $u^*(q)$ be the pleasure or utility derived from the bundle q, and let Y be the consumer's total income. The consumer's problem is

$$\text{Maximize } u^*(q) \quad \text{subject to } p'q \le Y \ .$$

The solution has the form $q(x)$ where $x = p/Y$. If one sets

$$g^*(x) = u^*[q(x)]$$

then the demand system $q(x)$ can be recovered by differentiation

$$q(x) = [x'(\partial/\partial x)g^*(x)]^{-1}(\partial/\partial x)g^*(x) \quad .$$

The recovery formula is called Roy's identity and the function $g^*(x)$ is called the consumer's indirect utility function. See Varian (1978) for regularity conditions and details.

These ideas may be adapted to emperical work by setting forth an indirect utility function $g(x,\lambda)$ which is thought to adequately approximate $g^*(x)$ over a region of interest $\mathfrak{X}$. Then Roy's identity is applied to obtain an approximating demand system. Usually one fits to consumer expenditure share data $q_i p_i/Y$ $i = 1, 2, \ldots, N$ although, as we have seen from Example 1, fitting $\ell n(q_i p_i/Y) - \ell n(q_N p_N/Y)$ to $\ell n[g_i(x)/g_N(x)]$ is preferable. The result is the expenditure system

$$q_i p_i/Y = f_i(x,\lambda) + e_i \qquad\qquad i = 1,2,\ldots,N-1$$

with

$$f_i(x,\lambda) = [x'(\partial/\partial x)g(x,\lambda)]^{-1} x_i(\partial/\partial x_i)g(x,\lambda) \quad .$$

The index i ranges to $N-1$ rather than N because expenditure shares sum to one for each consumer and the last share may be obtained by subtracting the rest. Converting to a vector notation, write

$$y = f(x,\lambda) + e$$

where y , $f(x,\lambda)$ and e are $N-1$ vectors. Measurements on n consumers yield the regression equations

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

Multivariate nonlinear least squares is often used to fit the data whence, referring to Notation 1, Chapter 3, the sample objective function is

$$s_n(\lambda) = (1/n)\textstyle\sum_{t=1}^n \frac{1}{2}[y_t - f(x_t,\lambda)]'(\hat{S}_n)^{-1}[y_t - f(x_t,\lambda)]$$

and

$$s(y,x,S,\lambda) = \tfrac{1}{2}[y - f(x,\lambda)]'S^{-1}[y - f(x,\lambda)]$$

where $\hat{S}$ is a preliminary estimator of $\mathcal{C}(e,e')$.

Suppose that the consumer's true indirect utility function is additive

$$q^*(x) = \Sigma_{i=1}^{N} g_i^*(x_i) \ .$$

Christensen, Jorgenson, and Lau (1975) have proposed that this supposition be tested by using a Translog indirect utility function

$$g(x,\lambda) = \Sigma_{i=1}^{N} \alpha_i \ln(x_i) + \Sigma_{i=1}^{N} \Sigma_{j=1}^{N} \beta_{ij} \ln(x_i) \ln(x_j)$$

to obtain the approximating expenditure system

$$f_i(x,\lambda) = \frac{\alpha_i + \Sigma_{j=1}^{N} \beta_{ij} \ln(x_j)}{-1 + \Sigma_{j=1}^{N} \beta_{Mj} \ln(x_j)}$$

with

$$\lambda = (\alpha_1, \alpha_2, \ldots, \alpha_{N-1}, \beta_{11}, \beta_{12}, \beta_{22}, \beta_{13}, \beta_{23}, \beta_{33}, \ldots, \beta_{1N}, \beta_{2N}, \ldots, \beta_{NN})'$$

and

$$\alpha_N = -1 - \Sigma_{j=1}^{N-1} \alpha_j \ , \ \beta_{ji} = \beta_{ij} \ \text{ for } i < j \ , \ \beta_{Mj} = \Sigma_{i=1}^{N} \beta_{ij} \ ,$$

then testing

$$H\colon \beta_{ij} = 0 \ \text{ for all } i \neq j \text{ against } A\colon \beta_{ij} \neq 0 \text{ for some } i \neq j \ .$$

This is a linear hypothesis of the form

$$h(\lambda) = H\lambda = 0$$

with

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

as a possible test statistic where $\hat{\lambda}_n$ minimizes $s_n(\lambda)$ and $\hat{V}$ is as defined in Section 5, Chapter 3.

The validity of this inference depends on whether a quadratic in logarithms is an adequate approximation to the consumer's true indirect utility function. For plausible alternative specifications of $g^*(x)$, it should be true that:

$$P(W > c) \doteq \alpha \quad \text{if } g^* \text{ is additive },$$

$$P(W > c) > \alpha \quad \text{if } g^* \text{ is not additive},$$

if the Translog specification is to be accepted as adequate. In this section we shall obtain an asymptotic approximation to $P(W > c)$ in order to shed some light on the quality of the approximation.

For an appropriately chosen sequence of N-vectors k_α, $\alpha = 1,2,3,\ldots$ the consumers indirect utility function must be of the Fourier form

$$g(x,\gamma) = u_o + b'x + \tfrac{1}{2}x'Cx$$
$$+ \Sigma_{\alpha=1}^{\infty}\{u_{o\alpha} + \Sigma_{j=1}^{\infty}[u_{j\alpha}\cos(jk_\alpha'x) - v_{j\alpha}\sin(jk_\alpha'x)]\}$$

where γ is vector of infinite length whose entries are b and some triangular arrangement of the $u_{j\alpha}$ and $v_{j\alpha}$; $C = -\Sigma_{\alpha=1}^{\infty}u_{o\alpha}k_\alpha k_\alpha'$. In consequence, the consumer's expenditure system $f(x,\gamma)$ is that which results by applying Roy's identity to $g(x,\gamma)$. The indirect utility function is additive if and only if the elementary N-vectors are the only vectors k_α which eneter $g(x,\gamma)$ with non-zero coefficients. That is, if and only if

$$g(x,\gamma) = u_o + b'x - \tfrac{1}{2}\Sigma_{\alpha=1}^{N}u_{o\alpha}x_\alpha^2$$
$$+ \Sigma_{\alpha=1}^{N}\{u_{o\alpha} + \Sigma_{j=1}^{\infty}[u_{j\alpha}\cos(jx_\alpha) - v_{j\alpha}\sin(jx_\alpha)]\} .$$

See Gallant (1981) for regularity conditions and details.

The situation is, then, as follows. The data is generated according to

$$y_t = f(x_t, \gamma^\circ) + e_t \qquad t = 1, 2, \ldots, n .$$

The fitted model is

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

with λ estimated by $\hat{\lambda}$ minimizing

$$s_n(\lambda) = (1/n)\Sigma_{t=1}^{n} s(y_t, x_t, \hat{S}_n, \lambda)$$

where

$$s(y, x, S, \lambda) = \tfrac{1}{2}[y - f(x, \lambda)]' S^{-1} [y - f(x, \lambda)] .$$

The probability $P[W > C]$ is to be approximated for plausible settings of the parameter γ° where

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

For simplicity, we shall compute power assuming that (y_t, x_t) are independently and identically distributed. Thus, $\mathcal{U}^*$ and $\mathcal{U}^\circ$ of Notations 2 and 3, Chapter are zero and the asymptotic distribution of W is the non-central chi-square. We assume that $\mathcal{E}(e) = 0$, $\mathcal{C}(ee') = \Sigma$, that $\hat{S}_n$ converges almost surely to Σ, and that $\sqrt{n}(\hat{S}_n - \Sigma)$ is bounded in probability. Direct computation using Notations 1 and 3 of Chapter 3 yields

Table 6. Data of Christensen, Jorgenson and Lau (1975).

	Durables		Non-durables		Services	
Year	Quantity	Price	Quantity	Price	Quantity	Price
1929	28.9645	33.9	98.1	38.4	96.1	31.6
1930	29.8164	32.2	93.5	36.4	89.5	32.1
1931	28.9645	31.4	93.1	31.1	84.3	30.9
1932	26.8821	23.9	85.9	26.5	77.1	28.8
1933	25.3676	31.3	82.9	26.8	76.8	26.1
1934	24.6104	27.7	88.5	30.2	76.3	26.8
1935	22.3387	28.8	93.2	31.5	79.5	26.8
1936	24.1371	32.9	103.8	31.6	83.8	27.2
1937	24.1371	29.0	107.7	32.7	86.5	28.3
1938	26.6928	28.4	109.3	31.1	83.7	29.1
1939	26.4088	30.5	115.1	30.5	86.1	29.2
1940	27.0714	29.4	119.9	30.9	88.7	29.5
1941	28.4912	28.9	127.6	33.6	91.8	30.8
1942	29.5325	31.7	129.9	39.1	95.5	32.4
1943	28.6806	38.0	134.0	43.7	100.1	34.2
1944	28.8699	37.7	139.4	46.2	102.7	36.1
1945	28.3966	39.0	150.3	47.8	106.3	37.3
1946	26.6928	44.0	158.9	52.1	116.7	38.9
1947	28.3966	65.3	154.8	58.7	120.8	41.7
1948	31.6149	60.4	155.0	62.3	124.6	44.4
1949	35.8744	50.4	157.4	60.3	126.4	46.1
1950	38.9980	59.2	161.8	60.7	132.8	47.4
1951	43.5414	60.0	165.3	65.8	137.1	49.9
1952	48.0849	64.2	171.2	66.6	140.8	52.6
1953	49.8833	57.5	175.7	66.3	145.5	55.4
1954	53.1016	68.3	177.0	66.6	150.4	57.2
1955	55.4680	63.5	185.4	66.3	157.5	58.5
1956	58.8756	62.2	191.5	67.3	164.8	60.2
1957	61.6206	56.5	194.8	69.4	170.3	62.2
1958	65.3122	66.7	196.8	71.0	175.8	64.2
1959	65.7854	63.3	205.0	71.4	184.7	66.0
1960	68.6251	73.1	208.2	72.6	192.3	68.0
1961	70.6129	72.1	211.9	73.3	200.0	69.1
1962	71.5594	72.4	218.5	73.9	208.7	70.4
1963	73.5472	72.5	223.0	74.9	217.6	71.7
1964	77.2387	76.3	233.3	75.8	229.7	72.8
1965	81.9715	82.3	244.0	77.3	240.7	74.3
1966	87.4615	84.3	255.5	80.1	251.6	76.5
1967	93.8981	81.0	259.5	81.9	264.0	78.8
1968	99.5774	81.0	270.2	85.3	275.0	82.0
1969	106.7710	94.4	276.4	89.4	287.2	86.1
1970	109.1380	85.0	282.7	93.6	297.3	90.5
1971	115.2900	88.5	287.5	96.6	306.3	95.8
1972	122.2000	100.0	299.3	100.0	322.4	100.0

Source: Gallant(1981)

$$\lambda^\circ \text{ minimizes } s_n^\circ(\lambda) = (1/n)\Sigma_{t=1}^n \tfrac{1}{2}\delta'(x_t,\lambda,\gamma^\circ)\ \Sigma^{-1}\delta(x_t,\lambda,\gamma^\circ) + (N-1)/2$$

$$\mathcal{I}^\circ = (1/n)\Sigma_{t=1}^n[(\partial/\partial\lambda')f(x_t,\lambda^\circ)]'\Sigma^{-1}[\Sigma + \delta(x_t,\lambda^\circ,\gamma^\circ)\delta'(x_t,\lambda^\circ,\gamma^\circ)]\Sigma^{-1}$$

$$\times\ [(\partial/\partial\lambda')f(x_t,\lambda^\circ)]$$

$$\mathcal{J}^\circ = (1/n)\Sigma_{t=1}^n[(\partial/\partial\lambda')f(x_t,\lambda^\circ)]'\Sigma^{-1}[(\partial/\partial\lambda')f(x_t,\lambda^\circ)]$$

$$- (1/n)\Sigma_{t=1}^n\ \Sigma_{i=1}^{N-1}\ \Sigma_{j=1}^{N-1}\ \delta_i(x_t,\lambda^\circ,\gamma^\circ)\Sigma^{ij}(\partial^2/\partial\lambda\partial\lambda')f_j(x_t,\lambda^\circ)$$

$$\alpha^\circ = n\lambda^{\circ\prime}H'(H V^\circ H')^{-1}H\lambda^\circ/2$$

$$V^\circ = (\mathcal{J}^\circ)^{-1}\mathcal{I}^\circ(\mathcal{J}^\circ)^{-1}$$

where

$$\delta(x_t,\lambda^\circ,\gamma) = f(x_t,\gamma^\circ) - f(x_t,\lambda^\circ)$$

and Σ^{ij} denotes the elements of Σ^{-1} .

Values of γ° were chosen as follows. The parameter γ° was truncated to vector of finite length by using only the multi-indices

$$k_\alpha = \begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\0\end{pmatrix}, \begin{pmatrix}1\\0\\1\end{pmatrix}, \begin{pmatrix}1\\1\\0\end{pmatrix}, \begin{pmatrix}0\\1\\1\end{pmatrix}, \begin{pmatrix}1\\1\\1\end{pmatrix}$$

and discarding the rest of the infinite sequence $\{k_\alpha\}_{\alpha=1}^\infty$. Let K denote the root sum of squares of the parameters of $g(x,\gamma)$ which are not associated with elementary vectors for k_α. For specified values of K, the parameters γ° were obtained by fitting $f(x,\gamma^\circ)$, subject to specified K, to the data used by Christenson, Jorgenson, and Lau (1975) which are shown in Table 6. This provides a sequence of indirect utility functions $g(x,\gamma^\circ)$ which increase in the degree of departure from additivity. When K = 0, $g(x,\gamma^\circ)$ is additive and when K is unconstrained the parameter γ° is free to adjust to the data as best it can.

The asymptotic approximation to $P(W \geq c)$ with c chosen to give a nominal .01 level test are shown in Table 7. For comparison, the power curve for W computed from the correct model - the Fourier expenditure system - is included in the table.

We see from Table 7 that the Translog test of explicit additivity is seriously flawed. The actual size of the test is much larger than the nominal significance level of .01 and the power curve is relatively flat. Power does increase near the null hypothesis, as one might expect, but it falls off again as departures from additivity become more extreme.

Table 7. Tests for an Additive Indirect Utility Function.

	Fourier		Translog	
K	Noncentrality	Power	Noncentrality	Power
0.0	0.0	0.010	8.9439	0.872
0.00046	0.0011935	0.010	8.9919	0.874
0.0021	0.029616	0.011	9.2014	0.884
0.0091	0.63795	0.023	10.287	0.924
0.033	4.6689	0.260	14.268	0.987
0.059	7.8947	0.552	15.710	0.993
0.084	82.875	1.000	13.875	0.984
unconstrained	328.61	1.000	10.230	0.922

8. REFERENCES

Aitchison,J. and S. M. Shen (1980) "Logistic-normal Distributions: Some Properties and Uses," Biometrica 67, 261-272.

Anderson, T. W. (1958), An Introduction to Multivariate Statistical Analysis. New York: John Wiley & Sons.

Blackorby, Charles, Daniel Primont, and R. Robert Russell (1978), Duality, Separability, and Functional Structure: Theory and Economic Applications. New York: North-Holland.

Caves, D. W. and Christensen, L. R. (1980), "Econometric Analysis of Residential Time-of-Use Electricity Experiments," Journal of Econometrics 14, 285-286.

Christensen, Laurits R., Dale W. Jorgenson, and Lawrence J. Lau (1975), "Transcendental Logarithmic Utility Functions," The American Economic Review 65, 367-383.

Deaton, Angus and John Muellbauer (1980), Economics and Consumer Behavior. Cambridge, England: Cambridge University Press.

Elbadawi, Ibrahim, A. Ronald Gallant, and Geraldo Souza (1983), "An Elasticity Can Be Estimated Consistently Without A Priori Knowledge of Functional Form," Econometrica 51, 1731-1751.

Gallant, A. Ronald, (1975) "Testing a Subset of the Parameters of a Nonlinear Regression Model", Journal of the American Statistical Association, 70, 927-932.

Gallant, A. Ronald (1981), "On the Bias in Flexible Functional Forms and an Essentially Unbiased Form: The Fourier Flexible Form," Journal of Econometrics 15, 211-245.

Gallant, A. Ronald (1982), "Unbiased Determination of Production Technologies," Journal of Econometrics 20, 285-323.

Gallant, A. Ronald and Alberto Holly (1980), "Statistical Inference in an Implicit, Nonlinear, Simultaneous Equation Model in the Context of Maximum Likelihood Estimation," Econometrica 48, 697-720.

Gallant, A. Ronald and Roger W. Koenker (1984), "Cost and Benefits of Peak-Load Pricing of Electricity: A Continuous-Time Econometric Approach," Journal of Econometrics ??, ???-???.

Gill, Philip E., Walter Murray and Margaret H. Wright (1981), Practical Optimization. New York: Academic Press.

Henderson, Harold V. and S. R. Searle (1979), "Vec and Vech Operators for Matrices, with some uses in Jacobians and Multivariate Statistics," The Canadian Journal of Statistics 7, 65-81.

Imhof, P. (1961), "Computing the Distribution of Quadratic Forms in Normal Variates," Biometrika 48, 419-426.

Rossi, Peter E. (1983), Specification and Analysis of Econometric Production Models, PhD. dissertation, University of Chicago, Chicago, Illinois.

Scheffe, Henry (1959), The Analysis of Variance, New York: John Wiley & Sons.

Snedecor, George W. and William G. Cochran (1980), Statistical Methods, Seventh Edition. Ames, Iowa: The Iowa State University Press.

Varian Hal R. (1978), Microeconomic Analysis. New York: W. W. Norton & Co.

Zellner, Arnold (1962), "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," Journal of the American Statistical Association 57, 348-368.

Rossi, Peter E. (1983), Specification and Analysis of Econometric Production Models, PhD. dissertation, University of Chicago, Chicago, Illinois.

Scheffe, Henry (1959), The Analysis of Variance, New York: John Wiley & Sons.

Snedecor, George W. and William G. Cochran (1980), Statistical Methods, Seventh Edition. Ames, Iowa: The Iowa State University Press.

Varian Hal R. (1978), Microeconomic Analysis. New York: W. W. Norton & Co.

Zellner, Arnold (1962), "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," Journal of the American Statistical Association 57, 348-368.

9. INDEX TO CHAPTER 6