THE INSTITUTE OF STATISTICS

UNIVERSITY OF NORTH CAROLINA SYSTEM

Nonlinear Statistical Models

by
A. Ronald Gallant

CHAPTER 8. NONLINEAR SIMULTANEOUS EQUATIONS MODELS

Institute of Statistics Mimeograph Series No. 1678

November 1985

NORTH CAROLINA STATE UNIVERSITY
Raleigh, North Carolina

NONLINEAR STATISTICAL MODELS

by

A. Ronald Gallant

Chapter 8. Nonlinear Simultaneous Equations Models

Copyright (C) 1985 by A. Ronald Gallant. All rights reserved.

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

	Anticipated Completion Date
1. Univariate Nonlinear Regression	Completed
1.0 Preface	
1.1 Introduction	
1.2 Taylor's Theorem and Matters of Notation	
1.3 Statistical Properties of Least Squares Estimators	
1.4 Methods of Computing Least Squares Estimators	
1.5 Hypothesis Testing	
1.6 Confidence Intervals	
1.7 References	
1.8 Index	
2. Univariate Nonlinear Regression: Special Situations	December 1985
3. A Unified Asymptotic Theory of Nonlinear Statistical Models	Completed
3.0 Preface	
3.1 Introduction	
3.2 The Data Generating Model and Limits of Cesaro Sums	
3.3 Least Mean Distance Estimators	
3.4 Method of Moments Estimators	
3.5 Tests of Hypotheses	
3.6 Alternative Representations of a Hypothesis	
3.7 Independently and Identically Distributed Regressors	
3.8 Constrained Estimation	
3.9 References	
3.10 Index	
4. Univariate Nonlinear Regression: Asymptotic Theory	Completed
4.0 Preface	
4.1 Introduction	
4.2 Regularity Conditions	
4.3 Characterizations of Least Squares Estimators and Test Statistics	
4.4 References	
4.5 Index	
5. Multivariate Linear Models: Review	Deleted

Table of Contents (continued)

	Anticipated Completion Date
6. Multivariate Nonlinear Models	Completed
6.0 Preface	
6.1 Introduction	
6.2 Least Squares Estimators and Matters of Notation	
6.3 Asymptotic Theory	
6.4 Hypothesis Testing	
6.5 Confidence Intervals	
6.6 Maximum Likelihood Estimators	
6.7 An Illustration of the Bias in Inference Caused by Misspecification	
6.8 References	
6.9 Index	
7. Linear Simultaneous Equations Models: Review	Deleted
8. Nonlinear Simultaneous Equations Models	Completed
8.0 Preface	
8.1 Introduction	
8.2 Three-Stage Least Squares	
8.3 The Dynamic Case: Generalized Method of Moments	
8.4 Hypothesis Testing	
8.5 Maximum Likelihood Estimation	
8.6 References	
8.7 Index	
9. Dynamic Nonlinear Models	Completed
9.0 Preface	
9.1 Introduction	
9.2 A Uniform Strong Law and a Central Limit Theorem for Dependent, Nonstationary Random Variables.	
9.3 Data Generating Process	
9.4 Least Mean Distance Estimators	
9.5 Method of Moments Estimators	
9.6 Hypothesis Testing	
9.7 References	
9.8 Index	

CHAPTER 8. Nonlinear Simultaneous Equations Models

In this chapter, we shall consider nonlinear, simultaneous equations models. These are multivariate models which cannot be written with the dependent variables equal to a vector valued function of the explanatory variables plus and additive error either because it is impossible or unnatural to do so; in short, the model is expressed in an implicit form $e = q(y,x,\theta)$ where e and y are vector valued. This is as much generality as is needed in applications. The model $q(e,y,x,\theta) = 0$ offers no more generality since the sort of regularity conditions that permit an asymptotic theory of inference also permit application of the implicit function theorem so that the form $e = q(y,x,\theta)$ must exit; the application where it cannot actually be produced is rare. In this rare instance, one substitutes numerical methods for the computation of e and its derivatives in the formulas that we shall derive.

There are two basic sets of statistical methods customarily employed with these models; those based on a method of moments approach with instrumental variables used to form the moment equations and those based on a maximum likelihood approach with some specific distribution specified for e. We shall discuss both approaches.

Frequently, these models are applied in situations where time indexes the observations and the vector of explanatory variables x_t has lagged values y_{t-1}, y_{t-2}, etc. of the dependent variable y_t as elements; models with a dynamic structure. In these situations, statistical methods cannot be derived from the asymptotic theory set forth in Chapter 3. But it is fairly easy to see intuitively, working by analogy with the statistical methods developed thus far from the asymptotic theory of Chapter 3, what the correct statistical procedures ought to be in dynamic models. Accordingly, we will lay down this intuitive base and develop the statistical methods for dynamic models in this chapter, deferring consideration of an asymptotic theory that will justify them to the next.

1. INTRODUCTION

In this chapter, the multivariate nonlinear regression model (Chapter 6) will be generalized in two ways.

First, we shall not insist that the model be written in explicit form where the dependent variables $y_{\alpha t}$ are solved out in terms of the independent variables x_t, the parameters θ_α, and the errors $e_{\alpha t}$. Rather, the model may be expressed in implicit form

$$q_\alpha(y_t, x_t, \theta_\alpha^o) = e_{\alpha t} \qquad t = 1, 2, \ldots, n; \quad \alpha = 1, 2, \ldots, M$$

where each $q_\alpha(y, x, \theta_\alpha)$ is a real-valued function, y_t is an L-vector, x_t is a k-vector, each θ_α^o is a p_α-dimensional vector of unknown parameters, and the $e_{\alpha t}$ represent unobservable observational or experimental errors. Note specifically that the number of equations M is not equal to the number of dependent variables L of necessity although in many applications this will be the case.

Secondly, the model can be dynamic which is to say that t indexes observations that are ordered in time and that the vector of independent variables x_t can include lagged values of the dependent variable (y_{t-1}, y_{t-2}, etc.) as elements. There is nothing in the theory (Chapter 9) that would preclude consideration of models of the form

$$q_{\alpha t}(y_t, \ldots, y_0, x_t, \ldots x_0, \theta_\alpha^o) = e_{\alpha t} \quad t = 1, 2, \ldots, n; \quad \alpha = 1, 2, \ldots, M$$

or similar schemes where the number of arguments of $q_{\alpha t}(\cdot)$ depends on t but they do not seem to arise in applications so we shall not consider them. If such a model is encountered, simply replace $q_\alpha(y_t, x_t, \theta_\alpha)$ by $q_{\alpha t}(y_t, \ldots, y_0, x_t, \ldots, x_0, \theta_\alpha)$ at every occurrence in Section 3 and thereafter.

Dynamic models frequently will have serially correlated errors ($\mathcal{C}(e_{\alpha s}, e_{\beta t}) \neq 0$ for $s \neq t$) and this fact will need to be taken into account in the analysis.

Two examples follow. The first has the classical regression structure, no lagged dependent variables, the second is dynamic.

EXAMPLE 1. (Consumer demand) This is a reformulation of Example 1 of Chapter 6. In Chapter 6, the analysis was conditional on prices and observed electricity expenditure whereas in theory it is preferable to condition on prices, income, and consumer demographic characteristics; that is, it is preferable to condition on the data in Table 1b and Table 1c of Chapter 6 rather than condition on Table 1b alone. In practice it is not clear that this is the case because the data of Tables 1a and 1b are of much higher quality than the data of Table 1c; there are several obvious errors in Table 1c such as a household with a dryer and no washer or a freezer and no refrigerator. Thus, it is not clear that we are not merely trading an errors in variables problem that arises from theory for a worse one that arises in practice.

To obtain the reformulated model, the data of Tables 1a, 1b, and 1c of Chapter 6 are transformed as follows.

$y_1 = \ell n$ (peak expenditure share) - ℓn (base expenditure share).

$y_2 = \ell n$ (intermediate expenditure share) - ℓn (base expenditure share),

$y_3 = \ell n$ (expenditure),

$r_1 = \ell n$ (peak price),

$r_2 = \ell n$ (intermediate price),

$r_3 = \ell n$ (base price),

$d_0 = 1$

$d_1 = \ell n$ [(10•peak price + 6•intermediate price + 8•base price)/24],

$d_2 = \ell n$ (income),

$d_3 = \ell n$ (residence size in SqFt),

d_4 = 1, if the residence is a duplex or apartment,
0, otherwise,

d_5 = 1, if the residence is a mobile home,
0, otherwise,

d_6 = ℓn (heat loss in Btuh), if the residence has central air conditioning,
0, otherwise,

d_7 = ℓn (window air Btuh), if the residence has window air conditioning,
0, otherwise,

d_8 = ℓn (number of household members + 1), if residence has an electric water heater,
0, otherwise,

d_9 = 1, if the residence has both an electric water heater and a washing machine,
0, otherwise,

d_{10} = ℓn (number of household members + 1), if residence has an electric dryer,
0, otherwise,

d_{11} = ℓn (refrigerator kw), if the residence has a refrigerator,
0, otherwise,

d_{12} = ℓn (freezer kw), if the residence has a freezer,
0, otherwise,

d_{13} = 1, if the residence has an electric range,
0, otherwise.

as notation, set

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}, \quad r = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix}, \quad d = \begin{bmatrix} d_0 \\ d_1 \\ \cdot \\ \cdot \\ \cdot \\ d_{13} \end{bmatrix}, \quad x = \begin{bmatrix} r \\ d \end{bmatrix}$$

$$y_t = \begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix}, \quad r_t = \begin{bmatrix} r_{1t} \\ r_{2t} \\ r_{3t} \end{bmatrix}, \quad d_t = \begin{bmatrix} d_{0t} \\ d_{1t} \\ \cdot \\ \cdot \\ \cdot \\ d_{13,t} \end{bmatrix}, \quad x_t = \begin{bmatrix} r_t \\ d_t \end{bmatrix}$$

These data are presumed to follow the model

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

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

$$y_{3t} = d_t'c + e_{3t}$$

where 1 denotes a vector of ones and

$$a = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}, \quad B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} = \begin{bmatrix} b_{(1)}' \\ b_{(2)}' \\ b_{(3)}' \end{bmatrix}, \quad c = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_3 \end{bmatrix}.$$

The matrix B is symmetric and $a_3 = -1$. With these conventions, the non-redundant parameters are

$$a_1, b_{11}, b_{12}, b_{13}, a_2, b_{22}, b_{23}, b_{33}, c_0, c_1, \ldots, c_{13} .$$

The errors

$$e_t = \begin{bmatrix} e_{1t} \\ e_{2t} \\ e_{3t} \end{bmatrix}$$

are taken to be independently and identically distributed each with zero mean.

The theory supporting this model was given in Chapter 6. The functional form of the third equation

$$y_{3t} = d_t'c + e_{3t}$$

and the variables entering into the equation were determined empirically from a data set of which Table 1 of Chapter 6 is a small subset; see Gallant and Koenker (1984).▌

EXAMPLE 2. (Intertemporal Consumption and Investment) The data shown in Table 1a is transformed as follows

$$y_t = \frac{\text{consumption at time t / population at time t}}{\text{consumption at time t - 1 / population at time t - 1}}$$

$$x_t = (1 + \text{stock returns at time t}) \frac{\text{deflator at time t - 1}}{\text{deflator at time t}} .$$

These data are presumed to follow the model

$$\beta^o (y_t)^{\alpha^o} x_t - 1 = e_t \qquad t = 1, 2, \ldots, 239.$$

Put

$$z_t = (1, y_{t-1}, x_{t-1})'.$$

What should be true of these data in theory is that

$$\mathcal{E}\,(e_t \otimes z_t) = 0, \qquad t = 2, 3, \ldots, 239,$$

$$\mathcal{E}\,(e_t \otimes z_t)(e_t \otimes z_t)' = \Sigma \qquad t = 2, 3, \ldots, 239,$$

$$\mathcal{E}\,(e_t \otimes z_t)(e_t \otimes z_t)' = 0 \qquad t \neq s.$$

Even though e_t is a scalar, we use the Kroneker product notation to keep the notation consistent with the later sections.

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 consumers problem is to allocate consumption and investment over time given a stream $\bar{w}_0, \bar{w}_1, \bar{w}_2, \ldots$ of incomes; $\bar{w}_t$ is the income that the consumer receives at time t. We suppose that the various consumption bundles available at time t can be mapped into a scalar quantity index c_t and that the consumer ranks various consumption streams $c_0, c_1, c_2, \ldots$ according to the utility indicator

$$U(\{c_t\}) = \sum_{t=1}^{\infty} \beta^t u(c_t)$$

where β is a discount factor, $0 < \beta < 1$, and $u(c)$ is a strictly concave increasing function. We suppose that there is a corresponding price index $\bar{p}_{0t}$ so that expenditure on consumption in each period can be computed according to $\bar{p}_{0t}c_t$. Above, we took c_t to be an index of consumption per capita on nondurables plus services and $\bar{p}_{0t}$ to be the corresponding implicit price deflator.

Further, suppose that the consumer has the choice of investing in a collection of N assets with maturities m_j, $j = 1, 2, \ldots, N$; asset j bought at time t cannot be sold until time $t + m_j$, or equivalently, an asset j bought at time $t - m_j$ cannot be sold until time t. Let q_{jt} denote the quantity of asset j held at time t, $\bar{p}_{jt}$ the price per unit of that asset at time t, and let $\bar{r}_{jt}$ denote the payoff at time t of asset j bought at time $t - m_j$. If, for example, the jth asset is a default-free, zero coupon bond with term to maturity m_j then $\bar{r}_{j,t+m_j}$ is the par value of the bond at time $t + m_j$; if the jth asset is a common stock then, definitionally, $m_j = 1$ and $\bar{r}_{jt} = \bar{p}_{jt} + \bar{d}_{jt}$ where $\bar{d}_{jt}$ is the dividend per share of the stock paid at time t, if any. Above, we took the first asset to be NYSE stocks weighted by value.

In Tables 1a and 1b, t is interpreted as the instant of time at the end of the month in which recorded. Nondurables and services, population, and the implicit deflator are assumed to be measured at the end of the month in which recorded, and assets are assumed to be purchased at the beginning of the month in which recorded. Thus, for a given row, nondurables and services divided by population is interpreted as c_t, the implicit price deflator is interpreted as $\bar{p}_{0t}$, and the return on asset j is interpreted as $(\bar{p}_{1t} + \bar{d}_{1t} - \bar{p}_{1t-1})/\bar{p}_{1t-1}$.

For example, if a three month bill is bought February 1 and sold April 30 it's return is recorded in the row for February. If $t+m_j$ refers to midnight April 30, the value of $t+m_j$ is recorded in the row for April and

$$\frac{c_{t+m_j}}{c_t} = \frac{\text{April Nondurables and Services / April Implicit Deflator}}{\text{January Nondurables and Services / January Implicit Deflator}}$$

$$\frac{r_{j,t+m_j}}{p_t} = (\text{February Return } +1)\,\frac{\text{January Implicit Deflator}}{\text{April Implicit Deflator}}\,.$$

As another, if a one month bill is bought April 1 and sold April 30 it is recorded in the row for April. If $t+m_j$ refers to midnight April 30 then

$$\frac{c_{t+m_j}}{c_t} = \frac{\text{April Nondurables and Services / April Implicit Deflator}}{\text{March Nondurables and Services / March Implicit Deflator}}$$

$$\frac{r_{j,t+m_j}}{p_t} = (\text{April Return } +1)\,\frac{\text{March Implicit Deflator}}{\text{April Implicit Deflator}}\,.$$

With these assumptions, the feasible consumption and investment plans must satisfy the sequence of budget constraints

$$\bar{p}_{0t}c_t + \sum_{t=1}^{N} \bar{p}_{jt}q_{jt} \le \bar{w}_t + \sum_{t=1}^{N} r_{jt}\bar{q}_{j,t-m_j},$$

$$0 \le q_{jt} .$$

The consumer seeks to maximize utility so the budget constraint is effectively

$$c_t = w_t + \sum_{t=1}^{N} r_{jt}q_{j,t-m_j} - \sum_{t=1}^{N} p_{jt}q_{jt}$$

$$0 \le q_{jt} ,$$

where $w_t = \bar{w}_t/\bar{p}_{0t}$, $r_{jt} = \bar{r}_{jt}/\bar{p}_{0t}$ and $p_{jt} = \bar{p}_{jt}/\bar{p}_{0t}$ or, in an obvious vector notation,

$$c_t = w_t + r_t'q_{t-m} - p_t'q_t,$$

$$0 \le q_t.$$

The sequences $\{w_t\}$, $\{r_t\}$, and $\{p_t\}$ are taken to be outside the consumers control or exogenously determined. Because these variables are outside the consumer's control, the sequence $\{c_t\}$ is determined by the budget constraint once $\{q_t\}$ is chosen. Thus, the only sequence that the consumer can control in attempting to maximize utility is the sequence $\{q_t\}$. The utility associated to some sequence $\{q_t\}$ is

$$V(\{q_t\}) = U(\{w_t + r_t'q_{t-m} - p_t'q_t\})$$

$$= \sum_{t=1}^{\infty} \beta^t u(w_t + r_t'q_{t-m} - p_t'q_t) .$$

We shall take the sequences $\{w_t\}$, $\{r_t\}$, and $\{p_t\}$ to be stochastic processes and shall assume that the consumer solves his optimization problem by choosing a sequence of functions Q_t of the form

$$Q_t(w_0, \ldots, w_t, r_0, \ldots, r_t, p_0, \ldots, p_t) \geq 0$$

which maximizes

$$\mathcal{E}_0 V(\{Q_t\}) = \mathcal{E}_0 \sum_{t=1}^{\infty} \beta^t u(w_t + r_t' Q_{t-m} - p_t' Q_t) .$$

where

$$\mathcal{E}_t(X) = \mathcal{E}[\ X \mid (w_0, \ldots, w_t, r_0, \ldots, r_t, p_0, \ldots, p_t)\ ] .$$

It may be that the consumer can achieve a higher expected utility by having regard to some sequence of vector-valued variables $\{v_t\}$ in addition to $\{w_t\}$, $\{r_t\}$, and $\{p_t\}$ in which case the argument list of Q and $\mathcal{E}$ above is replaced by the augmented argument list

$$(w_0, \ldots, w_t, r_0, \ldots, r_t, p_0, \ldots, p_t, v_0, \ldots, v_t).$$

Conceptually, v_t can be infinite dimensional because anything that is observable or knowable is admissible as additional information in improving the optimum. If one wishes to accommodate this possibility, let $\mathcal{B}_t$ be the smallest sigma-algebra such that the random variables

$$\{w_s, r_s, p_s, v_{sj}: s = 0, 1, \ldots, t;\ j = 1, 2, \ldots\}$$

are measurable, let Q_t be $\mathcal{B}_t$ measurable, and let $\mathcal{E}_t(X) = \mathcal{E}(X|\mathcal{B}_t)$. Either the augmented argument list or the sigma-algebra $\mathcal{B}_t$ is called the consumer's information set, depending on which approach is adopted. For our purposes, the augmented argument list provides enough generality.

Let Q_{0t} denote the solution to the consumer's optimization problem and consider the related problem

$$\text{maximize: } \mathcal{E}_s V(\{Q_t\}_{t=s}^{\infty}) = \mathcal{E}_s \sum_{t=s}^{\infty} \beta^t u(w_t + r_t' Q_{t-m} - p_t' Q_t) ,$$

$$\text{subject to: } Q_t = Q_t(w_0, \ldots, w_t, r_0, \ldots, r_t, p_0, \ldots, p_t, v_0, \ldots, v_t) \geq 0$$

with solution Q_{st}. Suppose we piece the two solutions together

$$\tilde{Q}_t = \begin{cases} Q_{0t} & 0 \leq t \leq s-1, \\ Q_{st} & s \leq t \leq \infty. \end{cases}$$

Then

$$\mathcal{E}_0 V(\{\tilde{Q}_t\}) = \mathcal{E}_0 \sum_{t=1}^{\infty} \beta^t u(w_t + r_t' \tilde{Q}_{t-m} - p_t' \tilde{Q}_t)$$

$$= \mathcal{E}_0 \sum_{t=1}^{s-1} \beta^t u(w_t + r_t' Q_{0t-m} - p_t' Q_{0t}) + \mathcal{E}_0 \mathcal{E}_s \sum_{t=s}^{\infty} \beta^t u(w_t + r_t' Q_{st-m} - p_t' Q_{st})$$

$$\geq \mathcal{E}_0 \sum_{t=1}^{s-1} \beta^t u(w_t + r_t' Q_{0t-m} - p_t' Q_{0t}) + \mathcal{E}_0 \mathcal{E}_s \sum_{t=s}^{\infty} \beta^t u(w_t + r_t' Q_{0t-m} - p_t' Q_{0t})$$

$$= \mathcal{E}_0 \sum_{t=1}^{\infty} \beta^t u(w_t + r_t' Q_{0t-m} - p_t' Q_{0t}) .$$

$$= \mathcal{E}_0 V(\{Q_{0t}\})$$

This inequality shows that Q_{0t} cannot be a solution to the consumer's optimization problem unless it is also a solution to the related problem for each s.

If we mechanically apply the Kuhn-Tucker theorem (Fiacco and McCormick, 1968) to the related problem

$$\text{maximize: } \mathcal{E}_s V(\{Q_t\}_{t=s}^{\infty}) = \mathcal{E}_s \Sigma_{t=s}^{\infty} \beta^t u(w_t + r_t' Q_{t-m} - p_t' Q_t \quad) ,$$

$$\text{subject to: } Q_t \geq 0$$

we obtain the first order conditions

$$(\partial/\partial Q_{js})[\mathcal{E}_0 V(\{Q_t\}) + \lambda_{js} Q_{js}] = 0,$$

$$\lambda_{js} Q_{js} \quad = 0,$$

$$\lambda_{js} \quad \leq 0.$$

where the λ_j are Lagrange multipliers. If a positive quantity of asset j is observed at time s -- if $Q_{js} > 0$ -- then we must have $\lambda_{js} = 0$ and

$$0 = (\partial/\partial Q_{js})\ \mathcal{E}_s \Sigma_{t=s}^{\infty}\ \beta^t u(w_t + r_t' Q_{t-m} - p_t' Q_t)$$

$$= (\partial/\partial Q_{js})\ \mathcal{E}_s\ \beta^s u(w_s + r_s' Q_{s-m} - p_s' Q_s)$$

$$+ (\partial/\partial Q_{js})\ \mathcal{E}_s\ \beta^{s+m_j} u(w_{s+m_j} + r'_{s+m_j} Q_{(s+m_j)-m} - p'_{s+m_j} Q_{s+m_j})$$

$$= - \beta^s\ (\partial/\partial c)\ u(c_s)\ p_{js} + \beta^{s+m_j}\ \mathcal{E}_s (\partial/\partial c)\ u(C_{s+m_j})\ r_{j,s+m_j}\ .$$

A little algebra reduces the first order conditions to

$$\beta^{m_j}\ \mathcal{E}_t[(\partial/\partial c)u(C_{t+m_j})/(\partial/\partial c)u(c_t)](r_{j,t+m_j}/p_{jt}) - 1 = 0$$

for $t = 0, 1, 2, \ldots$ and $j = 1, 2, \ldots, N$. See Lucas (1978) for a more rigorous derivation of these first order conditions.

Suppose that the consumer's preferences are given by some parametric function $u(c,\alpha)$ with $u(c) = u(c,\alpha^o)$ for some unknown value of α. Let $\gamma = (\alpha, \beta)$ and denote the true but unknown value of γ by γ^o. Let z_t be any vector-valued random variable whose value is known at time t and let the index j correspond to some sequence of securities $\{q_{jt}\}$ whose observed values are positive. Put

$$m_t(\gamma) = \left\{\beta^{m_j}[(\partial/\partial c)u(c_{t+m_j}, \alpha)/(\partial/\partial c)u(c_t, \alpha)](r_{j,t+m_j}/p_{jt}) - 1\right\} \otimes z_t .$$

If we could compute the moments of the sequence of random variables $\{m_t(\gamma^o)\}$ we could put

$$m_n(\gamma) = (1/n)\textstyle\sum_{t=1}^{n} m_t(\gamma)$$

and estimate γ^o by setting $m_n(\gamma) = \mathcal{E}m_n(\gamma^o)$ and solving for γ.

To this point, the operator $\mathcal{E}(\cdot)$ has represented expectations computed according to the consumer's subjective probability distribution. We shall now impose the Rational Expectations Hypothesis which states that the consumer's subjective probability distribution is the same distribution as the probability law governing the random variables $\{w_t\}$, $\{r_t\}$, $\{p_t\}$, and $\{v_t\}$. Under this assumption, the random variable $m_t(\gamma^o)$ will have first moment

$$\mathcal{E}\, m_t(\gamma^o)$$

$$= \mathcal{E}\, \mathcal{E}_t\left[\left\{(\beta^o)^{m_j}[(\partial/\partial c)u(c_{t+m_j}, \alpha^o)/(\partial/\partial c)u(c_t, \alpha^o)](r_{j,t+m_j}/p_{jt}) - 1\right\} \otimes z_t\right]$$

$$= \mathcal{E}\left[\mathcal{E}_t\left\{\beta^{o^{m_j}}[(\partial/\partial c)u(C_{t+m_j}, \alpha^o)/(\partial/\partial c)u(c_t, \alpha^o)](r_{j,t+m_j}/p_{jt}) - 1\right\} \otimes z_t\right]$$

$$= \mathcal{E}\,[0 \otimes z_t] = 0 \,.$$

For $s + m_j \leq t$ the random variables $m_t(\gamma^o)$ and $m_s(\gamma^o)$ are uncorrelated since

$$\mathcal{E}\, m_t(\gamma^o)\, m_s'(\gamma^o)$$

$$= \mathcal{E}\,\mathcal{E}_t\left[\left\{\beta^{o^{m_j}}[(\partial/\partial c)u(C_{t+m_j}, \alpha^o)/(\partial/\partial c)u(c_t, \alpha^o)](r_{j,t+m_j}/p_{jt}) - 1\right\} \otimes z_t\right]$$
$$\times \left[\left\{\beta^{o^{m_j}}[(\partial/\partial c)u(C_{s+m_j}, \alpha^o)/(\partial/\partial c)u(c_s, \alpha^o)](r_{j,s+m_j}/p_{js}) - 1\right\} \otimes z_s\right]'$$

$$= \mathcal{E}\left[\mathcal{E}_t\left\{\beta^{o^{m_j}}[(\partial/\partial c)u(C_{t+m_j}, \alpha^o)/(\partial/\partial c)u(c_t, \alpha^o)](r_{j,t+m_j}/p_{jt}) - 1\right\} \otimes z_t\right]$$
$$\times \left[\left\{\beta^{o^{m_j}}[(\partial/\partial c)u(c_{s+m_j}, \alpha^o)/(\partial/\partial c)u(c_s, \alpha^o)](r_{j,s+m_j}/p_{js}) - 1\right\} \otimes z_s\right]'$$

$$= \mathcal{E}\,[0 \otimes z_t]$$
$$\times \left[\left\{\beta^{o^{m_j}}[(\partial/\partial c)u(C_{s+m_j}, \alpha^o)/(\partial/\partial c)u(c_s, \alpha^o)](r_{j,s+m_j}/p_{js}) - 1\right\} \otimes z_s\right]'$$
$$= 0.$$

If we specialize these results to the utility function $u(c,\alpha) = c^{\alpha}/\alpha$ for $\alpha < 1$ and take q_{jt} to be common stocks we will have $u'(c,\alpha) = c^{\alpha}$, and $m_j = 1$ which gives the equations listed at the beginning of this discussion. ▮

Table 1a. Consumption and Stock Returns.

t	Year	Month	Nondurables and Services	Population	Value Weighted NYSE Returns	Implicit Deflator
0	1959	1	381.9	176.6850	0.0093695102	0.6818539
1	.	2	383.7	176.9050	0.0093310997	0.6823039
2	.	3	388.3	177.1460	0.0049904501	0.6814319
3	.	4	385.5	177.3650	0.0383739690	0.6830091
4	.	5	389.7	177.5910	0.0204769890	0.6846292
5	.	6	390.0	177.8300	0.0007165600	0.6876923
6	.	7	389.2	178.1010	0.0371922290	0.6893628
7	.	8	390.7	178.3760	-0.0113433900	0.6910673
8	.	9	393.6	178.6570	-0.0472779090	0.6930894
9	.	10	394.2	178.9210	0.0164727200	0.6945713
10	.	11	394.1	179.1530	0.0194594210	0.6950013
11	.	12	396.5	179.3860	0.0296911900	0.6958386
12	1960	1	396.8	179.5970	-0.0664901060	0.6960685
13	.	2	395.4	179.7880	0.0114439700	0.6967628
14	.	3	399.1	180.0070	-0.0114419700	0.6983212
15	.	4	404.2	180.2220	-0.0163223000	0.7013855
16	.	5	399.8	180.4440	0.0328373610	0.7016008
17	.	6	401.3	180.6710	0.0231378990	0.7024670
18	.	7	402.0	180.9450	-0.0210754290	0.7034826
19	.	8	400.4	181.2380	0.0296860300	0.7047952
20	.	9	400.2	181.5280	-0.0568203400	0.7061469
21	.	10	402.9	181.7960	-0.0045937700	0.7078680
22	.	11	403.8	182.0420	0.0472565590	0.7100049
23	.	12	401.6	182.2870	0.0478186380	0.7109064
24	1961	1	404.0	182.5200	0.0654135870	0.7106436
25	.	2	405.7	182.7420	0.0364446900	0.7108701
26	.	3	409.4	182.9920	0.0318523910	0.7105520
27	.	4	410.1	183.2170	0.0058811302	0.7100707
28	.	5	412.1	183.4520	0.0251120610	0.7100218
29	.	6	412.4	183.6910	-0.0296279510	0.7109602
30	.	7	410.4	183.9580	0.0303546690	0.7127193
31	.	8	411.5	184.2430	0.0251584590	0.7132442
32	.	9	413.7	184.5240	-0.0189705600	0.7138023
33	.	10	415.9	184.7830	0.0265103900	0.7136331
34	.	11	419.0	185.0160	0.0470347110	0.7136038
35	.	12	420.5	185.2420	0.0006585300	0.7143876
36	1962	1	420.8	185.4520	-0.0358958100	0.7155418
37	.	2	420.6	185.6500	0.0197925490	0.7177841
38	.	3	423.4	185.8740	-0.0055647301	0.7191781
39	.	4	424.8	186.0870	-0.0615112410	0.7203390
40	.	5	427.0	186.3140	-0.0834698080	0.7206089
41	.	6	425.2	186.5380	-0.0809235570	0.7208373
42	.	7	427.0	186.7900	0.0659098630	0.7203747
43	.	8	428.5	187.0580	0.0226466300	0.7218203
44	.	9	431.8	187.3230	-0.0487761200	0.7264938
45	.	10	431.0	187.5740	0.0039394898	0.7262181
46	.	11	433.6	187.7960	0.1114552000	0.7269373
47	.	12	434.1	188.0130	0.0139081300	0.7270214

(Continued next page)

Table 1a. (Continued).

t	Year	Month	Nondurables and Services	Population	Value Weighted NYSE Returns	Implicit Deflator
48	1963	1	434.7	188.2130	0.0508059190	0.7292386
49	.	2	433.7	188.3870	-0.0226500000	0.7299977
50	.	3	436.2	188.5800	0.0347222910	0.7297111
51	.	4	437.0	188.7900	0.0479895880	0.7295194
52	.	5	436.9	189.0180	0.0206833590	0.7308308
53	.	6	440.2	189.2420	-0.0178168900	0.7319400
54	.	7	442.1	189.4960	-0.0018435300	0.7335444
55	.	8	445.6	189.7610	0.0536292610	0.7349641
56	.	9	443.8	190.0280	-0.0126571200	0.7347905
57	.	10	444.2	190.2650	0.0286585090	0.7361549
58	.	11	445.8	190.4720	-0.0047020698	0.7375505
59	.	12	449.5	190.6680	0.0221940800	0.7385984
60	1964	1	450.1	190.8580	0.0256042290	0.7398356
61	.	2	453.7	191.0470	0.0181333610	0.7399162
62	.	3	456.6	191.2450	0.0173465290	0.7402540
63	.	4	456.7	191.4470	0.0051271599	0.7407488
64	.	5	462.1	191.6660	0.0166149310	0.7405324
65	.	6	463.8	191.8890	0.0158939310	0.7416990
66	.	7	466.0	192.1310	0.0202899800	0.7429185
67	.	8	468.5	192.3760	-0.0109651800	0.7432231
68	.	9	468.0	192.6310	0.0315313120	0.7448718
69	.	10	470.0	192.8470	0.0100951200	0.7457447
70	.	11	468.0	193.0390	0.0013465700	0.7465812
71	.	12	474.4	193.2230	0.0034312201	0.7476813
72	1965	1	474.5	193.3930	0.0377800500	0.7481560
73	.	2	477.4	193.5400	0.0066147698	0.7488479
74	.	3	474.5	193.7090	-0.0107356600	0.7506849
75	.	4	479.6	193.8880	0.0345496910	0.7525021
76	.	5	481.2	194.0870	-0.0047443998	0.7554032
77	.	6	479.5	194.3030	-0.0505878400	0.7593326
78	.	7	484.3	194.5280	0.0169978100	0.7602726
79	.	8	485.3	194.7610	0.0299301090	0.7601484
80	.	9	488.7	194.9970	0.0323472920	0.7605893
81	.	10	497.2	195.1950	0.0293272190	0.7626710
82	.	11	497.1	195.3720	0.0008636100	0.7648361
83	.	12	499.0	195.5390	0.0121703600	0.7671343
84	1966	1	500.1	195.6880	0.0100357400	0.7696461
85	.	2	501.5	195.8310	-0.0102875900	0.7730808
86	.	3	502.9	195.9990	-0.0215729900	0.7757009
87	.	4	505.8	196.1780	0.0233628400	0.7785686
88	.	5	504.8	196.3720	-0.0509349700	0.7793185
89	.	6	507.5	196.5600	-0.0109703900	0.7812808
90	.	7	510.9	196.7620	-0.0118703500	0.7827363
91	.	8	508.3	196.9840	-0.0748946070	0.7867401
92	.	9	510.2	197.2070	-0.0066132201	0.7894943
93	.	10	509.8	197.3980	0.0464050400	0.7910945
94	.	11	512.1	197.5720	0.0138342800	0.7922281
95	.	12	513.5	197.7360	0.0047225100	0.7933788

(Continued next page)

Table 1a. (Continued).

t	Year	Month	Nondurables and Services	Population	Value Weighted NYSE Returns	Implicit Deflator
96	1967	1	516.0	197.8920	0.0838221310	0.7941861
97	.	2	517.7	198.0370	0.0098125497	0.7948619
98	.	3	519.0	198.2060	0.0433843100	0.7959538
99	.	4	521.1	198.3630	0.0420965220	0.7965842
100	.	5	521.0	198.5370	-0.0415207000	0.7988484
101	.	6	523.1	198.7120	0.0232013710	0.8015676
102	.	7	522.1	198.9110	0.0482556600	0.8038690
103	.	8	525.5	199.1130	-0.0056581302	0.8058991
104	.	9	528.2	199.3110	0.0336121990	0.8076486
105	.	10	524.9	199.4980	-0.0276739710	0.8094875
106	.	11	527.9	199.6570	0.0078005102	0.8124645
107	.	12	531.9	199.8080	0.0307225010	0.8155668
108	1968	1	533.0	199.9200	-0.0389530290	0.8200750
109	.	2	533.9	200.0560	-0.0311505910	0.8231879
110	.	3	539.8	200.2080	0.0068653398	0.8262319
111	.	4	540.0	200.3610	0.0898963210	0.8285185
112	.	5	541.2	200.5360	0.0230161700	0.8318551
113	.	6	547.8	200.7060	0.0118528200	0.8335159
114	.	7	550.9	200.8980	-0.0211507900	0.8359049
115	.	8	552.4	201.0950	0.0163246690	0.8388849
116	.	9	551.0	201.2900	0.0422958400	0.8421053
117	.	10	552.1	201.4660	0.0111537600	0.8462235
118	.	11	556.7	201.6210	0.0562853110	0.8492905
119	.	12	554.1	201.7600	-0.0372401590	0.8521928
120	1969	1	557.0	201.8810	-0.0072337599	0.8560144
121	.	2	561.2	202.0230	-0.0502119700	0.8578047
122	.	3	560.6	202.1610	0.0314719300	0.8619336
123	.	4	561.9	202.3310	0.0213753300	0.8667023
124	.	5	566.5	202.5070	0.0029275999	0.8704325
125	.	6	563.9	202.6770	-0.0623450020	0.8751552
126	.	7	565.9	202.8770	-0.0630705430	0.8784238
127	.	8	569.4	203.0900	0.0504970810	0.8816298
128	.	9	568.2	203.3020	-0.0220447110	0.8856037
129	.	10	573.1	203.5000	0.0547974710	0.8888501
130	.	11	572.5	203.6750	-0.0314589110	0.8939738
131	.	12	572.4	203.8490	-0.0180749090	0.8981481
132	1970	1	577.2	204.0080	-0.0763489600	0.9019404
133	.	2	578.1	204.1560	0.0597185420	0.9058986
134	.	3	577.7	204.3350	-0.0023485899	0.9077376
135	.	4	577.1	204.5050	-0.0995834470	0.9123202
136	.	5	580.3	204.6920	-0.0611347710	0.9155609
137	.	6	582.0	204.8780	-0.0502832790	0.9176976
138	.	7	582.8	205.0860	0.0746088620	0.9208991
139	.	8	584.7	205.2940	0.0502020900	0.9235505
140	.	9	588.5	205.5070	0.0426676610	0.9276126
141	.	10	587.3	205.7070	-0.0160981810	0.9324025
142	.	11	587.6	205.8840	0.0521828200	0.9361811
143	.	12	592.6	206.0760	0.0617985200	0.9399258

(Continued next page)

Table 1a. (Continued).

t	Year	Month	Nondurables and Services	Population	Value Weighted NYSE Returns	Implicit Deflator
144	1971	1	592.2	206.2420	0.0492740680	0.9414049
145	.	2	594.5	206.3930	0.0149685600	0.9434819
146	.	3	592.4	206.5670	0.0441647210	0.9469953
147	.	4	596.1	206.7260	0.0341992900	0.9506794
148	.	5	596.3	206.8910	-0.0365711710	0.9548885
149	.	6	598.5	207.0530	0.0043891501	0.9597327
150	.	7	597.3	207.2370	-0.0398038400	0.9630002
151	.	8	599.1	207.4330	0.0409017500	0.9679519
152	.	9	601.1	207.6270	-0.0056930701	0.9698885
153	.	10	601.7	207.8000	-0.0395274310	0.9729101
154	.	11	604.9	207.9490	-0.0000956400	0.9752025
155	.	12	608.8	208.0880	0.0907427070	0.9797963
156	1972	1	607.9	208.1960	0.0241155400	0.9825629
157	.	2	610.3	208.3100	0.0308808110	0.9875471
158	.	3	618.9	208.4470	0.0091922097	0.9891743
159	.	4	620.6	208.5690	0.0066767102	0.9911376
160	.	5	622.3	208.7120	0.0176741590	0.9942150
161	.	6	623.7	208.8460	-0.0221355410	0.9961520
162	.	7	627.6	208.9880	-0.0018799200	0.9996813
163	.	8	629.7	209.1530	0.0382015890	1.0031760
164	.	9	631.7	209.3170	-0.0064313002	1.0079150
165	.	10	638.2	209.4570	0.0099495398	1.0117520
166	.	11	639.8	209.5840	0.0497901590	1.0151610
167	.	12	640.7	209.7110	0.0116306100	1.0190420
168	1973	1	643.4	209.8090	-0.0253177100	1.0247120
169	.	2	645.3	209.9050	-0.0398146990	1.0309930
170	.	3	643.3	210.0340	-0.0054550399	1.0399500
171	.	4	642.1	210.1540	-0.0464594700	1.0478120
172	.	5	643.2	210.2860	-0.0183557910	1.0541040
173	.	6	646.0	210.4100	-0.0088413004	1.0603720
174	.	7	651.9	210.5560	0.0521348010	1.0632000
175	.	8	643.4	210.7150	-0.0302029100	1.0792660
176	.	9	651.3	210.8630	0.0522540810	1.0815290
177	.	10	649.5	210.9840	-0.0018884100	1.0896070
178	.	11	651.3	211.0970	-0.1165516000	1.0993400
179	.	12	647.7	211.2070	0.0153318600	1.1093100
180	1974	1	648.4	211.3110	-0.0013036400	1.1215300
181	.	2	646.2	211.4110	0.0038444500	1.1363350
182	.	3	645.9	211.5220	-0.0243075400	1.1489390
183	.	4	648.6	211.6370	-0.0433935780	1.1558740
184	.	5	649.3	211.7720	-0.0352610800	1.1667950
185	.	6	650.3	211.9010	-0.0193944290	1.1737660
186	.	7	653.5	212.0510	-0.0730255170	1.1802600
187	.	8	654.5	212.2160	-0.0852853730	1.1926660
188	.	9	652.7	212.3830	-0.1098341000	1.2043820
189	.	10	654.5	212.5180	0.1671594000	1.2122230
190	.	11	651.2	212.6370	-0.0397416390	1.2205160
191	.	12	650.3	212.7480	-0.0234328400	1.2278950

(Continued next page)

Table 1a. (Continued).

t	Year	Month	Nondurables and Services	Population	Value Weighted NYSE Returns	Implicit Deflator
192	1975	1	653.7	212.8440	0.1358016000	1.2337460
193	.	2	657.4	212.9390	0.0607054380	1.2376030
194	.	3	659.4	213.0560	0.0293416310	1.2406730
195	.	4	659.7	213.1870	0.0470072290	1.2457180
196	.	5	670.4	213.3930	0.0546782990	1.2502980
197	.	6	669.7	213.5590	0.0517648310	1.2593700
198	.	7	668.3	213.7410	-0.0637501480	1.2721830
199	.	8	670.1	213.9000	-0.0203062710	1.2786150
200	.	9	670.2	214.0550	-0.0366309580	1.2821550
201	.	10	670.8	214.2000	0.0609995690	1.2904000
202	.	11	674.1	214.3210	0.0314961600	1.2966920
203	.	12	677.4	214.4460	-0.0105694800	1.3039560
204	1976	1	684.3	214.5610	0.1251743000	1.3081980
205	.	2	682.9	214.6550	0.0012425600	1.3069260
206	.	3	687.1	214.7620	0.0300192200	1.3092710
207	.	4	690.6	214.8810	-0.0108725300	1.3132060
208	.	5	688.7	215.0180	-0.0088088503	1.3206040
209	.	6	695.0	215.1520	0.0472505990	1.3256120
210	.	7	696.8	215.3110	-0.0073758499	1.3307980
211	.	8	699.6	215.4780	0.0005799900	1.3381930
212	.	9	702.5	215.6420	0.0261333100	1.3449110
213	.	10	705.6	215.7920	-0.0214380810	1.3520410
214	.	11	709.7	215.9240	0.0046152598	1.3587430
215	.	12	715.8	216.0570	0.0585772800	1.3657450
216	1977	1	717.6	216.1860	-0.0398427810	1.3720740
217	.	2	719.3	216.3000	-0.0162227190	1.3831500
218	.	3	716.5	216.4360	-0.0106509200	1.3884160
219	.	4	719.1	216.5650	0.0038957901	1.3953550
220	.	5	722.6	216.7120	-0.0126387400	1.4014670
221	.	6	721.5	216.8630	0.0509454310	1.4105340
222	.	7	728.3	217.0300	-0.0156951400	1.4159000
223	.	8	727.0	217.2070	-0.0140849800	1.4244840
224	.	9	729.1	217.3740	0.0006794800	1.4295710
225	.	10	735.7	217.5230	-0.0394544790	1.4350960
226	.	11	739.4	217.6550	0.0419719890	1.4442790
227	.	12	740.1	217.7850	0.0052549900	1.4508850
228	1978	1	738.0	217.8810	-0.0568409600	1.4581300
229	.	2	744.8	217.9870	-0.0121089800	1.4663000
230	.	3	750.5	218.1310	0.0318689010	1.4743500
231	.	4	750.4	218.2610	0.0833722430	1.4862740
232	.	5	750.3	218.4040	0.0186665390	1.5033990
233	.	6	753.1	218.5480	-0.0129163500	1.5146730
234	.	7	755.6	218.7200	0.0564879100	1.5199840
235	.	8	761.1	218.9090	0.0372171590	1.5284460
236	.	9	765.4	219.0780	-0.0063229799	1.5412860
237	.	10	765.2	219.2360	-0.1017461000	1.5541030
238	.	11	768.0	219.3840	0.0313147900	1.5640620
239	.	12	774.1	219.5300	0.0166718100	1.5694350

Source: Hansen and Singleton (1982,1984)

Table 1b. Treasury Bill Returns.

			Holding Period		
t	Year	Month	1 Month	3 Months	6 Months
0	1959	1	0.0021	0.0067620277	0.0149464610
1	.	2	0.0019	0.0067054033	0.0153553490
2	.	3	0.0022	0.0069413185	0.0156610010
3	.	4	0.0020	0.0071977377	0.0164365770
4	.	5	0.0022	0.0072308779	0.0162872080
5	.	6	0.0024	0.0076633692	0.0175679920
6	.	7	0.0025	0.0080889463	0.0190058950
7	.	8	0.0019	0.0075789690	0.0191299920
8	.	9	0.0031	0.0097180605	0.0230187180
9	.	10	0.0030	0.0103986260	0.0247714520
10	.	11	0.0026	0.0101703410	0.0219579940
11	.	12	0.0034	0.0112402440	0.0246732230
12	1960	1	0.0053	0.0111309290	0.0253483060
13	.	2	0.0029	0.0101277830	0.0230150220
14	.	3	0.0035	0.0106238130	0.0225163700
15	.	4	0.0019	0.0076926947	0.0172802210
16	.	5	0.0027	0.0076853037	0.0174299480
17	.	6	0.0024	0.0079696178	0.0172280070
18	.	7	0.0013	0.0055410862	0.0131897930
19	.	8	0.0017	0.0055702925	0.0127488370
20	.	9	0.0016	0.0057605505	0.0142772200
21	.	10	0.0022	0.0058845282	0.0143817660
22	.	11	0.0013	0.0053367615	0.0121047500
23	.	12	0.0016	0.0060379505	0.0138461590
24	1961	1	0.0019	0.0055921078	0.0121918920
25	.	2	0.0014	0.0058113337	0.0127416850
26	.	3	0.0020	0.0065517426	0.0141991380
27	.	4	0.0017	0.0060653687	0.0131639240
28	.	5	0.0018	0.0057235956	0.0120902060
29	.	6	0.0020	0.0059211254	0.0130860810
30	.	7	0.0018	0.0057165623	0.0123114590
31	.	8	0.0014	0.0056213140	0.0128748420
32	.	9	0.0017	0.0059502125	0.0135532620
33	.	10	0.0019	0.0056616068	0.0136462450
34	.	11	0.0015	0.0057950020	0.0132193570
35	.	12	0.0019	0.0064492226	0.0142288210
36	1962	1	0.0024	0.0067474842	0.0148699280
37	.	2	0.0020	0.0068224669	0.0148422720
38	.	3	0.0020	0.0068684816	0.0148569350
39	.	4	0.0022	0.0069818497	0.0147231820
40	.	5	0.0024	0.0068957806	0.0145040750
41	.	6	0.0020	0.0068334341	0.0141508580
42	.	7	0.0027	0.0073847771	0.0149892570
43	.	8	0.0023	0.0072803497	0.0155196190
44	.	9	0.0021	0.0071101189	0.0151845220
45	.	10	0.0025	0.0069708824	0.0148054360
46	.	11	0.0020	0.0068755150	0.0143437390
47	.	12	0.0023	0.0072675943	0.0150616170

(Continued next page)

Table 1b. (Continued).

			Holding Period		
t	Year	Month	1 Month	3 Months	6 Months
48	1963	1	0.0025	0.0073993206	0.0151528120
49	.	2	0.0023	0.0074235201	0.0153222080
50	.	3	0.0023	0.0073300600	0.0150877240
51	.	4	0.0025	0.0073734522	0.0152205230
52	.	5	0.0024	0.0073573589	0.0153070690
53	.	6	0.0023	0.0076377392	0.0158174040
54	.	7	0.0027	0.0075825453	0.0157427790
55	.	8	0.0025	0.0083107948	0.0174001460
56	.	9	0.0027	0.0086047649	0.0178933140
57	.	10	0.0029	0.0085899830	0.0179922580
58	.	11	0.0027	0.0088618994	0.0184588430
59	.	12	0.0029	0.0088895559	0.0187247990
60	1964	1	0.0030	0.0089538097	0.0187361240
61	.	2	0.0026	0.0088821650	0.0185148720
62	.	3	0.0031	0.0091063976	0.0192459820
63	.	4	0.0029	0.0089648962	0.0188522340
64	.	5	0.0026	0.0087850094	0.0184588430
65	.	6	0.0030	0.0088362694	0.0184062720
66	.	7	0.0030	0.0087610483	0.0180916790
67	.	8	0.0028	0.0088040829	0.0182579760
68	.	9	0.0028	0.0087461472	0.0181269650
69	.	10	0.0029	0.0090366602	0.0189046860
70	.	11	0.0029	0.0089949369	0.0189571380
71	.	12	0.0031	0.0096279383	0.0203764440
72	1965	1	0.0028	0.0097374916	0.0201922660
73	.	2	0.0030	0.0097503662	0.0203632120
74	.	3	0.0036	0.0101563930	0.0207239390
75	.	4	0.0031	0.0098274946	0.0205006600
76	.	5	0.0031	0.0099694729	0.0204553600
77	.	6	0.0035	0.0098533630	0.0202448370
78	.	7	0.0031	0.0096805096	0.0197677610
79	.	8	0.0033	0.0096729994	0.0199424030
80	.	9	0.0031	0.0096987486	0.0202448370
81	.	10	0.0031	0.0102273230	0.0216147900
82	.	11	0.0035	0.0102949140	0.0214961770
83	.	12	0.0033	0.0104292630	0.0217467550
84	1966	1	0.0038	0.0114099980	0.0241273640
85	.	2	0.0035	0.0117748980	0.0235883000
86	.	3	0.0038	0.0116740470	0.0245782140
87	.	4	0.0034	0.0115190740	0.0244190690
88	.	5	0.0041	0.0118079190	0.0243724580
89	.	6	0.0038	0.0116353030	0.0239523650
90	.	7	0.0035	0.0116611720	0.0240478520
91	.	8	0.0041	0.0120524170	0.0255184170
92	.	9	0.0040	0.0125415330	0.0282398460
93	.	10	0.0045	0.0136051180	0.0286009310
94	.	11	0.0040	0.0133793350	0.0278792380
95	.	12	0.0040	0.0131639240	0.0271854400

(Continued next page)

Table 1b. (Continued).

t	Year	Month	Holding Period: 1 Month	3 Months	6 Months
96	1967	1	0.0043	0.0122823720	0.0254547600
97	.	2	0.0036	0.0114994050	0.0231827500
98	.	3	0.0039	0.0115102530	0.0232950450
99	.	4	0.0032	0.0102145670	0.0208503010
100	.	5	0.0033	0.0094506741	0.0197293760
101	.	6	0.0027	0.0087846518	0.0192024710
102	.	7	0.0031	0.0100209710	0.0223804710
103	.	8	0.0031	0.0105757710	0.0241363050
104	.	9	0.0032	0.0111052990	0.0248435740
105	.	10	0.0039	0.0111955400	0.0259439950
106	.	11	0.0036	0.0115144250	0.0258800980
107	.	12	0.0033	0.0126451250	0.0286339520
108	1968	1	0.0040	0.0127922300	0.0286719800
109	.	2	0.0039	0.0123600960	0.0252685550
110	.	3	0.0038	0.0128250120	0.0269453530
111	.	4	0.0043	0.0130600930	0.0273784400
112	.	5	0.0045	0.0138714310	0.0293196440
113	.	6	0.0043	0.0146449800	0.0306105610
114	.	7	0.0048	0.0135078430	0.0283122060
115	.	8	0.0042	0.0131379370	0.0273113250
116	.	9	0.0043	0.0132676360	0.0269986390
117	.	10	0.0044	0.0130808350	0.0272654290
118	.	11	0.0042	0.0140209200	0.0285472870
119	.	12	0.0043	0.0139169690	0.0287613870
120	1969	1	0.0053	0.0156394240	0.0327807660
121	.	2	0.0046	0.0158472060	0.0329054590
122	.	3	0.0046	0.0159218310	0.0330774780
123	.	4	0.0053	0.0152308940	0.0315673350
124	.	5	0.0048	0.0150020120	0.0310289860
125	.	6	0.0051	0.0155957940	0.0331374410
126	.	7	0.0053	0.0159739260	0.0351030830
127	.	8	0.0050	0.0181180240	0.0373669860
128	.	9	0.0062	0.0177775620	0.0368773940
129	.	10	0.0060	0.0182124380	0.0373669860
130	.	11	0.0052	0.0178167820	0.0382032390
131	.	12	0.0064	0.0192197560	0.0406687260
132	1970	1	0.0060	0.0201528070	0.0415455100
133	.	2	0.0062	0.0201845170	0.0399575230
134	.	3	0.0057	0.0175420050	0.0353578330
135	.	4	0.0050	0.0160522460	0.0327900650
136	.	5	0.0053	0.0176727770	0.0373940470
137	.	6	0.0058	0.0176465510	0.0366872550
138	.	7	0.0052	0.0163348910	0.0338231330
139	.	8	0.0053	0.0161305670	0.0333294870
140	.	9	0.0054	0.0157260890	0.0328800680
141	.	10	0.0046	0.0148160460	0.0328979490
142	.	11	0.0046	0.0148533580	0.0315511230
143	.	12	0.0042	0.0125738380	0.0254547600

(Continued next page)

Table 1b. (Continued).

t	Year	Month	Holding Period: 1 Month	3 Months	6 Months
144	1971	1	0.0038	0.0123342280	0.0249764920
145	.	2	0.0033	0.0104724170	0.0215225220
146	.	3	0.0030	0.0085597038	0.0180971620
147	.	4	0.0028	0.0086690187	0.0190058950
148	.	5	0.0029	0.0100854640	0.0215620990
149	.	6	0.0037	0.0109888320	0.0232796670
150	.	7	0.0040	0.0131714340	0.0277613400
151	.	8	0.0047	0.0134450200	0.0293101070
152	.	9	0.0037	0.0109888320	0.0238093140
153	.	10	0.0037	0.0116611720	0.0252420900
154	.	11	0.0037	0.0109502080	0.0227504970
155	.	12	0.0037	0.0108597280	0.0224862100
156	1972	1	0.0029	0.0091836452	0.0203764440
157	.	2	0.0025	0.0084762573	0.0186685320
158	.	3	0.0027	0.0085406303	0.0192197560
159	.	4	0.0029	0.0096729994	0.0224862100
160	.	5	0.0030	0.0091643333	0.0205764770
161	.	6	0.0029	0.0096213818	0.0212192540
162	.	7	0.0031	0.0102015730	0.0230324270
163	.	8	0.0029	0.0093638897	0.0218169690
164	.	9	0.0034	0.0115317110	0.0257205960
165	.	10	0.0040	0.0114973780	0.0267322060
166	.	11	0.0037	0.0118292570	0.0260753630
167	.	12	0.0037	0.0123860840	0.0266788010
168	1973	1	0.0044	0.0129822490	0.0277190210
169	.	2	0.0041	0.0144283770	0.0300337080
170	.	3	0.0046	0.0148830410	0.0311747790
171	.	4	0.0052	0.0163263080	0.0353578330
172	.	5	0.0051	0.0159218310	0.0385923390
173	.	6	0.0051	0.0176727770	0.0363070960
174	.	7	0.0064	0.0192872290	0.0398482080
175	.	8	0.0070	0.0212606190	0.0439169410
176	.	9	0.0068	0.0221503970	0.0455567840
177	.	10	0.0065	0.0179196600	0.0393607620
178	.	11	0.0056	0.0188183780	0.0385221240
179	.	12	0.0064	0.0187247990	0.0404498580
180	1974	1	0.0063	0.0191034080	0.0378948450
181	.	2	0.0058	0.0192459820	0.0377205610
182	.	3	0.0056	0.0191277270	0.0383199450
183	.	4	0.0075	0.0215357540	0.0431109670
184	.	5	0.0075	0.0226182940	0.0460462570
185	.	6	0.0060	0.0207901000	0.0436338190
186	.	7	0.0070	0.0189310310	0.0415712590
187	.	8	0.0060	0.0196549890	0.0430125000
188	.	9	0.0081	0.0232796670	0.0497723820
189	.	10	0.0051	0.0157541040	0.0382287500
190	.	11	0.0054	0.0199817420	0.0407782790
191	.	12	0.0070	0.0190396310	0.0393673180

(Continued next page)

Table 1b. (Continued).

t	Year	Month	Holding Period 1 Month	3 Months	6 Months
192	1975	1	0.0058	0.0180318360	0.0361442570
193	.	2	0.0043	0.0144213440	0.0303153990
194	.	3	0.0041	0.0137765410	0.0290288930
195	.	4	0.0044	0.0141247510	0.0309060810
196	.	5	0.0044	0.0138258930	0.0305682420
197	.	6	0.0041	0.0132120850	0.0280662780
198	.	7	0.0048	0.0151025060	0.0319100620
199	.	8	0.0048	0.0159218310	0.0347890850
200	.	9	0.0053	0.0162349940	0.0358686450
201	.	10	0.0056	0.0166865590	0.0402698520
202	.	11	0.0041	0.0141117570	0.0294467210
203	.	12	0.0048	0.0141172410	0.0310940740
204	1976	1	0.0047	0.0131934880	0.0284404750
205	.	2	0.0034	0.0120286940	0.0256940130
206	.	3	0.0040	0.0127384660	0.0284404750
207	.	4	0.0042	0.0125932690	0.0277762410
208	.	5	0.0037	0.0125379560	0.0275914670
209	.	6	0.0043	0.0140357020	0.0306911470
210	.	7	0.0047	0.0136423110	0.0298480990
211	.	8	0.0042	0.0132082700	0.0285162930
212	.	9	0.0044	0.0129737850	0.0276085140
213	.	10	0.0041	0.0129562620	0.0273720030
214	.	11	0.0040	0.0124713180	0.0261569020
215	.	12	0.0040	0.0112657550	0.0235443120
216	1977	1	0.0036	0.0110250710	0.0232796670
217	.	2	0.0035	0.0120649340	0.0258269310
218	.	3	0.0038	0.0119923350	0.0256673100
219	.	4	0.0038	0.0115835670	0.0246578450
220	.	5	0.0037	0.0119640830	0.0254743100
221	.	6	0.0040	0.0127894880	0.0269986390
222	.	7	0.0042	0.0127229690	0.0268921850
223	.	8	0.0044	0.0137610440	0.0295736790
224	.	9	0.0043	0.0142158270	0.0304151770
225	.	10	0.0049	0.0150877240	0.0321971180
226	.	11	0.0050	0.0159218310	0.0339560510
227	.	12	0.0049	0.0154412980	0.0331405400
228	1978	1	0.0049	0.0156871080	0.0336534980
229	.	2	0.0046	0.0164700750	0.0349206920
230	.	3	0.0053	0.0164960620	0.0349899530
231	.	4	0.0054	0.0166230200	0.0351681710
232	.	5	0.0051	0.0163058040	0.0359042880
233	.	6	0.0054	0.0170561080	0.0374757050
234	.	7	0.0056	0.0180656910	0.0388967990
235	.	8	0.0056	0.0175942180	0.0384653810
236	.	9	0.0062	0.0193772320	0.0403403040
237	.	10	0.0068	0.0206396580	0.0438541170
238	.	11	0.0070	0.0223728420	0.0483353140
239	.	12	0.0078	0.0232532020	0.0490025280

Source: Hansen and Singleton (1982, 1984)

2. THREE-STAGE LEAST SQUARES.

Multivariate responses y_t, L-vectors, are assumed to be determined by k-dimensional independent variables x_t according to the system of simultaneous equations

$$q_\alpha(y_t, x_t, \theta_\alpha^o) = e_{\alpha t} \qquad \alpha = 1, 2, \ldots, L; \quad t = 1, 2, \ldots, n,$$

where each $q_\alpha(y,x,\theta_\alpha)$ is a real-valued function, each θ_α^o is a p_α-dimensional vector of unknown parameters, and the $e_{\alpha t}$ represent unobservable observational or experimental errors. The analysis is conditional on the sequence of independent variables $\{x_t\}$ as described in Section 2 of Chapter 3 and the x_t do not contain lagged values of the y_t as elements. See the next section for the case when they do.

For any set of values $e_1, e_2, \ldots, e_L$ of the errors, any admissible value for the vector x of independent variables, and any admissible value of the parameters $\theta_1, \theta_2, \ldots, \theta_m$, the system of equations

$$q_\alpha(y, x, \theta_\alpha) = e_\alpha \qquad \alpha = 1, 2, \ldots, L$$

is assumed to determine y uniquely; if the equations have multiple roots, there is some rule for determining which solution is meant. Moreover, the solution must be a continuous function of the errors, the parameters, and the independent variables. However, one is not obligated to actually be able to compute y given these variables or even to have complete knowledge of the system in order to use the methods described below it is just that the theory (Chapter 3) on which the methods are based relies on this assumption for its validity.

There is one feature of implicit models that one should not lose sight of in applications. If an equation of the system

$$q_\alpha(y, x, \theta_\alpha) = e_\alpha$$

is transformed using some one-to-one function $\Psi(e)$ to obtain

$$\Psi[q_\alpha(y, x, \theta_\alpha)] = u_\alpha$$

where

$$u_\alpha = \Psi[e_\alpha],$$

the result is still a nonlinear equation in implicit form which is equivalent to the original equation for the purpose of determining y from knowledge of the independent variables, parameters, and errors. However, setting aside the identity transformation, the distribution of the random variable u_α will differ from that of e_α. Thus one has complete freedom to use transformations of this sort in applications an attempt to make the error distribution more nearly normally distributed. Doing so is sound statistical practice.

In an application it may be the case that not all the equations of the system are known or it may be that one is simply not interested in some of them. Reorder the equations as necessary so that it is the first M of the L equations above that are of interest, let θ be a p-vector that consists of the non-redundant parameters in the set θ_1, θ_2, ..., θ_M, and let

$$q(y,x,\theta) = \begin{bmatrix} q_1(y,x,\theta_1) \\ q_2(y,x,\theta_2) \\ \vdots \\ q_M(y,x,\theta_M) \end{bmatrix}, \qquad e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_M \end{bmatrix}$$

$$q(y_t, x_t, \theta) = \begin{bmatrix} q_1(y_t, x_t, \theta_1) \\ q_2(y_t, x_t, \theta_2) \\ \vdots \\ q_M(y_t, x_t, \theta_M) \end{bmatrix}, \qquad e_t = \begin{bmatrix} e_{1,t} \\ e_{2,t} \\ \vdots \\ e_{M,t} \end{bmatrix}$$

We assume that the error vectors e_t are independently and identically distributed with mean zero and unknown variance-covariance matrix Σ,

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

Independence implies a lack of correlation, viz.

$$\mathcal{C}(e_t, e_s') = 0, \qquad t \neq s.$$

This is the grouped-by-observation or multivariate arrangement of the data with the equation index α thought of as the fastest moving index and the observation index t the slowest. The alternative arrangement is the grouped-by-equation ordering with t the fastest moving index and α the slowest. As we saw in Chapter 6, the multivariate scheme has two advantages, it facilitates writing code and it meshes better with the underlying theory (Chapter 3). However, the grouped-by-equation formulation is more prevalent in the literature because it was the dominant form in the linear simultaneous equations literature and got carried over when the nonlinear literature developed. We shall develop the ideas using the multivariate scheme and then conclude with a summary in the alternative notation. Let us illustrate with the first example.

EXAMPLE 1. (continued). Recall that the model is

$$y_{1t} = \ln\left[(a_1 + r_t'b_{(1)} - y_{3t}1'b_{(1)})/(a_3 + r_t'b_{(3)} - y_{3t}1'b_{(3)})\right] + e_{1t}$$

$$y_{2t} = \ln\left[(a_2 + r_t'b_{(2)} - y_{3t}1'b_{(2)})/(a_3 + r_t'b_{(3)} - y_{3t}1'b_{(3)})\right] + e_{2t}$$

$$y_{3t} = d_t'c + e_{3t}$$

where 1 denotes a vector of ones,

$$y_t = \begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix}, \quad r_t = \begin{bmatrix} r_{1t} \\ r_{2t} \\ r_{3t} \end{bmatrix}, \quad d_t = \begin{bmatrix} d_{0t} \\ d_{1t} \\ \vdots \\ d_{13,t} \end{bmatrix}, \quad x_t = \begin{bmatrix} r_t \\ d_t \end{bmatrix}$$

and

$$a = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}, \quad B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} = \begin{bmatrix} b_{(1)}' \\ b_{(2)}' \\ b_{(3)}' \end{bmatrix}, \quad c = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_3 \end{bmatrix}.$$

The matrix B is symmetric and $a_3 = -1$.

Our interest centers in the first and second equations so we write

$$q(y,x,\theta) = \begin{bmatrix} y_1 - \ln \dfrac{\theta_1 + \theta_2 \cdot r_1 + \theta_3 \cdot r_2 + \theta_4 \cdot r_3 - (\theta_2 + \theta_3 + \theta_4) \cdot y_3}{-1 + \theta_4 \cdot r_1 + \theta_7 \cdot r_2 + \theta_8 \cdot r_3 - (\theta_4 + \theta_7 + \theta_8) \cdot y_3} \\ \\ y_2 - \ln \dfrac{\theta_5 + \theta_3 \cdot r_1 + \theta_6 \cdot r_2 + \theta_7 \cdot r_3 - (\theta_3 + \theta_6 + \theta_7) \cdot y_3}{-1 + \theta_4 \cdot r_1 + \theta_7 \cdot r_2 + \theta_8 \cdot r_3 - (\theta_4 + \theta_7 + \theta_8) \cdot y_3} \end{bmatrix}$$

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

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

Taking values from Table 1 of Chapter 6, we have

$$x_1 = (r_1' \mid d_1')'$$

$$= (1.36098,\ 1.05082,\ 0.058269 \mid 1,\ 0.99078,\ 9.7410,\ 6.39693,\ 0,\ 1,\ 0.0000,\ 9.4727,\ 1.09861,\ 1,\ 0.00000,\ -0.35667,\ 0.00000,\ 0)'$$

$$x_2 = (1.36098,\ 1.05082,\ 0.058269 \mid 1,\ 0.99078,\ 9.5104,\ 6.80239,\ 0,\ 0,\ 0.0000,\ 0.0000,\ 1.94591,\ 1,\ 0.00000,\ -0.35667,\ 0.27763,\ 1)'$$

•
•
•

$$x_{19} = (1.36098,\ 1.05082,\ 0.058269 \mid 1,\ 0.99078,\ 8.7903,\ 6.86693,\ 0,\ 0,\ 0.0000, 10.0858,\ 1.38629,\ 1,\ 0.00000,\ 0.58501,\ 0.27763,\ 1)'$$

$$x_{20} = (1.36098,\ 1.05082,\ 0.576613 \mid 1,\ 1.07614,\ 9.1050,\ 6.64379,\ 0,\ 0,\ 0.0000,\ 0.0000,\ 1.60944,\ 1,\ 1.60944,\ 0.58501,\ 0.00000,\ 1)'$$

•
•
•

$$x_{40} = (1.36098,\ 1.05082,\ 0.576613 \mid 1,\ 1.07614,\ 11.1664,\ 7.67415,\ 0,\ 0,\ 9.7143,\ 0.0000,\ 1.60944,\ 1,\ 1.60944,\ 0.33647,\ 0.27763,\ 1)'$$

$$x_{41} = (1.36098,\ 1.36098,\ 0.058269 \mid 1,\ 1.08293,\ 8.8537,\ 6.72383,\ 0,\ 0,\ 8.3701,\ 0.0000,\ 1.09861,\ 1,\ 1.09861,\ 0.58501,\ 0.68562,\ 1)'$$

•
•
•

$$x_{224} = (1.88099,\ 1.36098,\ 0.576613 \mid 1,\ 1.45900,\ 8.8537,\ 6.88653,\ 0,\ 0,\ 0.0000,\ 9.2103,\ 0.69315,\ 1,\ 0.69315,\ 0.58501,\ 0.00000,\ 1)'$$

for the independent or exogenous variables and we have

$$y_1 = (2.45829,\ 1.59783,\ -0.7565)'$$

$$y_2 = (1.82933,\ 0.89091,\ -0.2289)'$$

$$\vdots$$

$$y_{19} = (2.33247,\ 1.31287,\ 0.3160)'$$

$$y_{20} = (1.84809,\ 0.86533,\ -0.0751)'$$

$$\vdots$$

$$y_{40} = (1.32811,\ 0.72482,\ 0.9282)'$$

$$y_{41} = (2.18752,\ 0.90133,\ 0.1375)'$$

$$\vdots$$

$$y_{224} = (1.06851,\ 0.51366,\ 0.1475)'$$

for the dependent or endogenous variables.

One might ask why we are handling the model in this way rather than simply substituting $d'c$ for y_3 above and then applying the methods of Chapter 6. After all, the theory on which we rely is nonstochastic and we just tacked on an error term at a convenient moment in the discussion. As to the theory, it would have been just as defensible to substitute $d'c$ for y_3 in the nonstochastic phase of the analysis and then tack on the error term. By way of reply, the approach we are taking seems to follow the natural progression of ideas. Throughout Chapter 6, the variable y_3 was regarded as being a potentially error ridden proxy for what we really had in mind. Now, a direct remedy seems more in order than a complete reformulation of the model. Moreover, the specification $\mathcal{E}y_3 = d'c$ was data determined and is rather ad hoc. It is probably best just to rely on it for the purpose of suggesting instrumental variables and not to risk the specification error a substitution of $d'c$ for y_3 might entail. ▌

Three-stage least squares is a method of moments type estimator where instrumental variables are used to form the moment equations. That is, letting

z_t denote some K-vector of random variables, one forms sample moments

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

where

$$m(y_t, x_t, \theta) = q(y_t, x_t, \theta) \otimes z_t = \underset{M \cdot K \qquad\qquad\qquad 1}{\begin{bmatrix} q_1(y_t, x_t, \theta_1) \cdot z_t \\ q_2(y_t, x_t, \theta_2) \cdot z_t \\ \vdots \\ q_M(y_t, x_t, \theta_M) \cdot z_t \end{bmatrix}},$$

equates them to population moments

$$m_n(\theta) = \mathcal{E}[m_n(\theta^o)],$$

and uses the solution $\hat{\theta}$ as the estimate of θ^o. If, is as usually the case, the dimension $M \cdot K$ of $m_n(\theta)$ exceeds the dimension p of θ these equations will not have a solution. In this case, one applies the generalized least squares heuristic and estimates θ^o by that value $\hat{\theta}$ that minimizes

$$S(\theta, V) = [n \cdot m_n(\theta)]' \; V^{-1} [n \cdot m_n(\theta)]$$

with

$$V = \mathcal{C}\left\{[n \cdot m_n(\theta^o)],\; [n \cdot m_n(\theta^o)]'\right\}.$$

To apply these ideas, one must compute $\mathcal{E}[m_n(\theta^o)]$ and $\mathcal{C}[m_n(\theta^o), m_n'(\theta^o)]$. Obviously there is an incentive to make this computation as easy as possible. Since

$$m_n(\theta^o) = (1/n)\sum_{t=1}^{n} e_t \otimes z_t,$$

we will have $\mathcal{E}[m_n(\theta^o)] = 0$ if z_t is uncorrelated with e_t and

$$\mathcal{C}\left\{[n\cdot m_n(\theta^o)],\ [n\cdot m_n(\theta^o)]'\right\} = \sum_{t=1}^{n}(\Sigma \otimes z_t z_t') \quad = \quad \Sigma \otimes \sum_{t=1}^{n} z_t z_t'$$

if $\{z_t\}$ is independent of $\{e_t\}$. These conditions will obtain (Problem 1) if we impose the requirement that

$$z_t = Z(x_t)$$

where $Z(x)$ is some (possibly nonlinear) function of the independent variables.

We shall also want z_t to be correlated with $q(y_t, x_t, \theta)$ for values of θ other than θ^o or the method will be vacuous (Problem 2). This last condition is made plausible by the requirement that $z_t = Z(x_t)$ but, strictly speaking, direct verification of the identification condition (Chapter 3, Section 4)

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

is required. This is an almost sure limit. Its computation is discussed in Section 2 of Chapter 3 but it usually suffices to check that

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

As we remarked in Chapter 1, few are going to take the trouble to verify this condition in an application but it is prudent to be on guard for violations that are easily detected (Problem 3).

The matrix V is unknown so we adopt the same strategy that was used in multivariate least squares: Form a preliminary estimate $\hat{\theta}^{\#}$ of θ^o and then estimate V from residuals. Let

$$\hat{\theta}^{\#} = \text{argmin}_{\theta}\ s(\theta,\ I \otimes \sum_{t=1}^{n} z_t z_t' \ )$$

and put

$$\hat{V} = \left[(1/n)\sum_{t=1}^{n} q(y_t,\ x_t,\ \hat{\theta}^{\#})\ q'(y_t,\ x_t,\ \hat{\theta}^{\#})\right] \otimes \left[\sum_{t=1}^{n} z_t z_t'\right].$$

There are two alternative estimators of V in the literature. The first of these affords some protection against heteroskedasticity

$$\bar{V} = \sum_{t=1}^{n}\left[q(y_t,x_t,\hat{\theta}^{\#}) \otimes z_t\right]\left[q(y_t,x_t,\hat{\theta}^{\#}) \otimes z_t\right]'.$$

The second uses two-stage least-squares residuals to estimate Σ. The two-stage least-squares estimate of the parameters of the single equation

$$q_\alpha(y_t,x_t,\theta_\alpha) = e_{\alpha t} \qquad t = 1,\ 2,\ \ldots,\ n$$

is

$$\hat{\theta}_\alpha^{\#} = \operatorname{argmin}_\theta \left[\sum_{t=1}^{n} q_\alpha(y_t,x_t,\theta_\alpha)\cdot z_t\right]' \left[\sum_{t=1}^{n} z_t z_t'\right]^{-1} \left[\sum_{t=1}^{n} q_\alpha(y_t,x_t,\theta_\alpha)\cdot z_t\right].$$

Two-stage least squares is vestigial terminology left over from the linear case (Problem 4). Letting

$$\bar{\sigma}_{\alpha\beta} = (1/n)\sum_{t=1}^{n} q_\alpha(y_t,x_t,\hat{\theta}_\alpha^{\#})\cdot q_\beta(y_t,x_t,\hat{\theta}_\beta^{\#}),$$

the estimate of Σ is the matrix $\bar{\Sigma}$ with typical element $\bar{\sigma}_{\alpha\beta}$, viz.

$$\bar{\Sigma} = [\bar{\sigma}_{\alpha\beta}],$$

and the estimate of V is

$$\bar{V} = \bar{\Sigma} \otimes \left[\sum_{t=1}^{n} z_t z_t'\right].$$

Suppose that one worked by analogy with the generalized least squares approach used in Chapter 6 and viewed

$$[y - f(\theta)] = \sum_{t=1}^{n} q(y_t,x_t,\theta) \otimes z_t$$

as a nonlinear regression in vector form and viewed

$$S(\theta,\hat{V}) = [y - f(\theta)]' \hat{V}^{-1}[y - f(\theta)]$$

$$= [\Sigma_{t=1}^{n} q(y_t,x_t,\theta) \otimes z_t]'\hat{V}^{-1}[\Sigma_{t=1}^{n} q(y_t,x_t,\theta) \otimes z_t]$$

as the objective function for the generalized least-squares estimator of θ. One would conclude that the estimated variance-covariance matrix of $\hat{\theta}$ was

$$\hat{C} = \left[\{(\partial/\partial\theta')[y-f(\hat{\theta})]\}' \; V^{-1}\{(\partial/\partial\theta')[y-f(\hat{\theta})]\}\right]^{-1}$$

$$= \left\{[\Sigma_{t=1}^{n}(\partial/\partial\theta')q(y_t,x_t,\hat{\theta}) \otimes z_t]'\hat{V}^{-1}[\Sigma_{t=1}^{n}(\partial/\partial\theta')q(y_t,x_t,\hat{\theta}) \otimes z_t]\right\}^{-1}$$

$$= \left\{[\Sigma_{t=1}^{n}Q(y_t,x_t,\hat{\theta}) \otimes z_t]'\hat{V}^{-1}[\Sigma_{t=1}^{n}Q(y_t,x_t,\hat{\theta}) \otimes z_t]\right\}^{-1}.$$

This intuitive approach does lead to the correct answer (Problem 5).

The Gauss-Newton correction vector can be deduced in this way as well (Problem 6),

$$D(\theta,V) = -\left\{[\Sigma_{t=1}^{n}Q(y_t,x_t,\theta) \otimes z_t]'V^{-1}[\Sigma_{t=1}^{n}Q(y_t,x_t,\theta) \otimes z_t]\right\}^{-1}$$

$$\times \left\{[\Sigma_{t=1}^{n}Q(y_t,x_t,\theta) \otimes z_t]'V^{-1}[\Sigma_{t=1}^{n}q(y_t,x_t,\theta) \otimes z_t]\right\}$$

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

0) Choose a starting estimate θ_0. Compute $D_0 = D(\theta_0,V)$ and find a λ_0 between zero and one such that $S(\theta_0+\lambda_0 D_0,V) < S(\theta_0,V)$.

1) Let $\theta_1 = \theta_0 + \lambda_0 D_0$. Compute $D_1 = D(\theta_1,V)$ and find a λ_1 between zero and one such that $S(\theta_1+\lambda_1 D_1,V) < S(\theta_1,V)$.

2) Let $\theta_2 = \theta_1 + \lambda_1 D_1$.

.
.
.

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

In summary, the three-stage least-squares estimator is computed as follows. The set of equations of interest are

$$q(y_t,x_t,\theta) = \begin{bmatrix} q_1(y_t,x_t,\theta_1) \\ q_2(y_t,x_t,\theta_2) \\ \vdots \\ q_M(y_t,x_t,\theta_M) \end{bmatrix} \qquad t = 1, 2, \ldots, n$$

and instrumental variables of the form

$$z_t = Z(x_t)$$

are selected. The objective function that defines the estimator is

$$S(\theta,V) = [\Sigma_{t=1}^{n} q(y_t,x_t,\theta) \otimes z_t]' \; V^{-1} \; [\Sigma_{t=1}^{n} q(y_t,x_t,\theta) \otimes z_t].$$

One minimizes $S(\theta,\ I \otimes \Sigma_{t=1}^{n} z_t z_t')$ to obtain a preliminary estimate $\hat{\theta}^{\#}$, viz.

$$\hat{\theta}^{\#} = \operatorname{argmin}_{\theta} S(\theta,\ I \otimes \Sigma_{t=1}^{n} z_t z_t'),$$

and puts

$$\hat{V} = \left[(1/n)\Sigma_{t=1}^{n} q(y_t,x_t,\hat{\theta}^{\#})\ q'(y_t,x_t,\hat{\theta}^{\#})\right] \otimes \left[\Sigma_{t=1}^{n} z_t\ z_t'\right].$$

The estimate of θ^{o} is the minimizer $\hat{\theta}$ of $S(\theta,\hat{V})$, viz.

$$\hat{\theta} = \operatorname{argmin}_{\theta} S(\theta,\hat{V}).$$

The estimated variance-covariance of $\hat{\theta}$ is

$$\hat{C} = \left\{[\Sigma_{t=1}^{n} Q(y_t,x_t,\hat{\theta}) \otimes z_t]'\hat{V}^{-1}[\Sigma_{t=1}^{n} Q(y_t,x_t,\hat{\theta}) \otimes z_t]\right\}^{-1}$$

where

$$Q(y_t, x_t, \theta) = (\partial/\partial\theta')q(y_t, x_t, \theta) \ .$$

We illustrate with the example.

EXAMPLE 1 (continued). A three-stage least-squares fit of the model

$$q(y,x,\theta) = \begin{bmatrix} y_1 - \ell n \dfrac{\theta_1 + \theta_2 \cdot r_1 + \theta_3 \cdot r_2 + \theta_4 \cdot r_3 - (\theta_2 + \theta_3 + \theta_4) \cdot y_3}{-1 + \theta_4 \cdot r_1 + \theta_7 \cdot r_2 + \theta_8 \cdot r_3 - (\theta_4 + \theta_7 + \theta_8) \cdot y_3} \\ \\ y_2 - \ell n \dfrac{\theta_5 + \theta_3 \cdot r_1 + \theta_6 \cdot r_2 + \theta_7 \cdot r_3 - (\theta_3 + \theta_6 + \theta_7) \cdot y_3}{-1 + \theta_4 \cdot r_1 + \theta_7 \cdot r_2 + \theta_8 \cdot r_3 - (\theta_4 + \theta_7 + \theta_8) \cdot y_3} \end{bmatrix}$$

to the data of Table 1, Chapter 6, is shown as Figure 1.

Figure 1. Example 1 Fitted by Nonlinear Three-Stage Least Squares.

SAS Statements:

```
PROC MODEL OUT=MOD01;
ENDOGENOUS Y1 Y2 Y3;
EXOGENOUS  R1 R2 R3 D0 D1 D2 D3 D4 D5 D6 D7 D8 D9 D10 D11 D12 D13;
PARMS T1 T2 T3 T4 T5 T6 T7 T8 C0 C1 C2 C3 C4 C5 C6 C7 C8 C9 C10 C11 C12 C13;
PEAK= T1+T2*R1+T3*R2+T4*R3-(T2+T3+T4)*Y3;
INTER=T5+T3*R1+T6*R2+T7*R3-(T3+T6+T7)*Y3;
BASE= -1+T4*R1+T7*R2+T8*R3-(T4+T7+T8)*Y3;
Y1=LOG(PEAK/BASE);  Y2=LOG(INTER/BASE);
Y3=D0*C0+D1*C1+D2*C2+D3*C3+D4*C4+D5*C5+D6*C6+D7*C7+D8*C8+D9*C9+D10*C10+D11*C11
   +D12*C12+D13*C13;
PROC SYSNLIN DATA=EG01 MODEL=MOD01 N3SLS METHOD=GAUSS MAXIT=50 CONVERGE=1.E-8
             SDATA=IDENTITY OUTS=SHAT OUTEST=THAT;
INSTRUMENTS  R1 R2 R3 D0 D1 D2 D3 D4 D5 D6 D7 D8 D9 D10 D11 D12 D13 / NOINT;
FIT Y1 Y2  START=(T1 -2.98  T2 -1.16  T3 0.787  T4 0.353  T5 -1.51  T6 -1.00
             T7 0.054  T8 -0.474);
PROC SYSNLIN DATA=EG01 MODEL=MOD01 N3SLS METHOD=GAUSS MAXIT=50 CONVERGE=1.E-8
             SDATA=SHAT ESTDATA=THAT;
INSTRUMENTS  R1 R2 R3 D0 D1 D2 D3 D4 D5 D6 D7 D8 D9 D10 D11 D12 D13 / NOINT;
```

Output:

SAS 7

SYSNLIN PROCEDURE

NONLINEAR 3SLS PARAMETER ESTIMATES

PARAMETER	ESTIMATE	APPROX. STD ERROR	'T' RATIO	APPROX. PROB>!T!	1ST STAGE R-SQUARE
T1	-2.13788	0.58954	-3.63	0.0004	0.6274
T2	-1.98939	0.75921	-2.62	0.0094	0.5473
T3	0.70939	0.15657	4.53	0.0001	0.7405
T4	0.33663	0.05095	6.61	0.0001	0.7127
T5	-1.40200	0.15226	-9.21	0.0001	0.7005
T6	-1.13890	0.18429	-6.18	0.0001	0.5225
T7	0.02913	0.04560	0.64	0.5236	0.5468
T8	-0.50050	0.04517	-11.08	0.0001	0.4646

NUMBER OF OBSERVATIONS		STATISTICS FOR SYSTEM	
USED	220	OBJECTIVE	0.15893
MISSING	4	OBJECTIVE*N	34.96403

COVARIANCE OF RESIDUALS MATRIX USED FOR ESTIMATION

S	Y1	Y2
Y1	0.17159	0.09675
Y2	0.09675	0.09545

We obtain

$$\hat{\theta} = \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \hat{\theta}_3 \\ \hat{\theta}_4 \\ \hat{\theta}_5 \\ \hat{\theta}_6 \\ \hat{\theta}_7 \\ \hat{\theta}_8 \end{bmatrix} = \begin{bmatrix} -2.13788 \\ -1.98939 \\ 0.70939 \\ 0.33663 \\ -1.40200 \\ -1.13890 \\ 0.02913 \\ -0.50050 \end{bmatrix} \qquad \text{(from Figure 1)},$$

$$\hat{\Sigma} = \begin{bmatrix} 0.17159 & 0.09675 \\ 0.09675 & 0.09545 \end{bmatrix} \qquad \text{(from Figure 1)},$$

$$S(\hat{\theta},\hat{\Sigma}) = 34.96403 \qquad \text{(from Figure 1)}.$$

These estimates are little changed from the multivariate least-squares estimates

$$\hat{\theta} = \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \hat{\theta}_3 \\ \hat{\theta}_4 \\ \hat{\theta}_5 \\ \hat{\theta}_6 \\ \hat{\theta}_7 \\ \hat{\theta}_8 \end{bmatrix} = \begin{bmatrix} -2.92458 \\ -1.28675 \\ 0.81857 \\ 0.36116 \\ -1.53759 \\ -1.04896 \\ 0.03009 \\ -0.46742 \end{bmatrix} \qquad \text{(from Figure 3c, Chapter 6)}.$$

The main impact of the use of three-stage least squares has been to inflate the estimated standard errors. ▌

The alternative notational convention is obtained by combining all the observations pertaining to a single equation into an n-vector

$$q_\alpha(\theta_\alpha) = \begin{bmatrix} q_\alpha(y_1, x, \theta_\alpha) \\ q_\alpha(y_2, x, \theta_\alpha) \\ \vdots \\ q_\alpha(y_n, x, \theta_\alpha) \end{bmatrix}_{n \quad 1} \qquad (\alpha = 1, 2, \ldots, M)$$

and then stacking these vectors equation-by-equation to obtain

$$q(\theta) = \begin{bmatrix} q_1(\theta_1) \\ q_2(\theta_2) \\ \vdots \\ q_M(\theta_M) \end{bmatrix}_{n \cdot M \quad 1}$$

with

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_M \end{bmatrix}_{\sum_{t=1}^{M} p_\alpha \quad 1} .$$

If desired, one can impose across equations restrictions by deleting the redundant entries of θ. Arrange the instrumental variables into a matrix Z as follows

$$Z = \begin{bmatrix} z_1' \\ z_2' \\ \vdots \\ z_n' \end{bmatrix}_{n \quad K}$$

and put

$$P_Z = Z\,(Z'Z)^{-1}Z' \quad .$$

With these conventions the three-stage least-squares objective function is

$$S[\theta,(\Sigma \otimes Z'Z)] = q'(\theta)(\Sigma^{-1} \otimes P_Z)q(\theta).$$

An estimate of Σ can be obtained by computing either

$$\hat{\theta}^{\#} = \operatorname{argmin}_{\theta} S[\theta,(I \otimes Z'Z)]$$

or

$$\hat{\theta}^{\#}_{\alpha} = \operatorname{argmin}_{\theta_\alpha} q'_{\alpha}(\theta_{\alpha})\; P_Z\; q_{\alpha}(\theta_{\alpha}) \qquad (\alpha = 1,\ 2,\ \ldots,\ M)$$

and letting $\hat{\Sigma}$ be the matrix with typical element

$$\hat{\sigma}_{\alpha\beta} = (1/n)q'_{\alpha}(\hat{\theta}^{\#})q_{\alpha}(\hat{\theta}^{\#}_{\alpha}).$$

The estimate of θ^{o} is

$$\hat{\theta} = \operatorname{argmin}_{\theta} S[\theta,(\hat{\Sigma} \otimes Z'Z)]$$

with estimated variance-covariance matrix

$$\hat{C} = \left\{\left[(\partial/\partial\theta')q(\hat{\theta})\right]'(\hat{\Sigma}^{-1} \otimes P_Z)\left[(\partial/\partial\theta')q(\hat{\theta})\right]\right\}^{-1}.$$

These expressions are svelte by comparison with the summation notation used used above. But the price of beauty is an obligation to assume that the errors e_t each have the same variance Σ and are uncorrelated with the sequence $\{z_t\}$. Neither the correction for heteroskedasticity suggested above nor the correction for autocorrelation discussed in the next section can be accommodated within this notational framework.

Amemiya (1977) considered the question of the optimal choice of instrumental variables and found that the optimal choice is obtained if the columns of Z span the same space as the union of the spaces spanned by the columns of $\mathcal{E}(\partial/\partial\theta_\alpha')q(\theta_\alpha^o)$. This can necessitate a large number of columns of Z which presumably adds to the small sample variance of the estimator but will have no effect asymptotically. He proposes some alternative three-stage least-squares type estimators obtained by replacing $(\Sigma^{-1} \otimes P_Z)$ with a matrix that has smaller rank but achieves the same asymptotic variance. He also shows that the three-stage least-squares is not as efficient asymptotically as the maximum likelihood estimator, discussed in Section 5.

The most disturbing aspect of three-stage least squares estimators is that they are not invariant to the choice of instrumental variables. Various sets of instrumental variables can lead to quite different parameter estimates even though the model specification and data remain the same. A dramatic illustration of this point can be had by looking at the estimates published by Hansen and Singleton (1982,1984). Bear in mind when looking at their results that their maximum likelihood estimator is obtained by assuming a distribution for the data and then imposing parametric restrictions implied by the model rather than deriving the likelihood implied by the model and an assumed error distribution; Section 5 takes the latter approach as does Amemiya's comparison.

One would look to results on the optimal choice of instrumental variables for some guidance that would lead to a resolution of this lack of invariance problem. But they do not provide it. Leaving aside the issue of either having to know the parameter values or estimate them, one would have to specify the error distribution in order to compute $\mathcal{E}(\partial/\partial\theta_\alpha')q(\theta_\alpha^o)$. But if the error distribution is known, maximum likelihood is the method of choice.

In practice, the most common approach is to use the independent variables x_{it} and low order monomials in x_{it} such as $(x_{it})^2$ or $(x_{it}x_{jt})$ as instrumental variables, making no attempt to find the most efficient set using the results on efficiency. We shall return to this issue at the end of the next section.

PROBLEMS

1. Consider the system of nonlinear equations

$$a_0^o + a_1^o \ln y_{1t} + a_2^o x_t = e_{1t},$$

$$t = 1, 2, \ldots$$

$$b_0^o + b_1^o y_{1t} + y_{2t} + b_2^o x_t = e_{2t}$$

where the errors $e_t = (e_{1t}, e_{2t})$ are normally distributed and the independent variable x_t follows the replication pattern

$$x_t = 0, 1, 2, 3, 0, 1, 2, 3, \ldots .$$

Put

$$\theta = (b_0, b_1, b_2)',$$

$$m(y_t, x_t, \theta) = (b_0 + b_1 y_{1t} + y_{2t} + b_2 x_t) \otimes z_t ,$$

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

$$z_t = (1, x_t, x_t^2)' .$$

Show that

$$\lim_{n\to\infty} m_n(\theta^o) = \mathcal{E}\, m_n(\theta^o) = 0 , \qquad \text{almost surely,}$$

$$\lim_{n\to\infty} m_n(\theta) = \begin{bmatrix} 1 & c\sum_{x=0}^3 \exp(-a_2^o x) & 1.5 \\ 1.5 & c\sum_{x=0}^3 x\exp(-a_2^o x) & 3.5 \\ 3.5 & c\sum_{x=0}^3 x^2\exp(-a_2^o x) & 9 \end{bmatrix} \begin{bmatrix} b_0 - b_0^o \\ b_1 - b_1^o \\ b_2 - b_2^o \end{bmatrix}$$

almost surely where $c = (1/4)\exp[\mathrm{Var}(e_1)/2 - a_0^o]$.

2. Referring to Problem 1, show that if $a_2 \neq 0$ then

$$\ell im_{n\to\infty} m_n(\theta) = 0 \quad \Rightarrow \quad \theta^o = \theta.$$

3. Referring to Problem 1, show that the model is not identified if either $a_2 = 0$ or $z_t = (1, x_t)$.

4. Consider the linear system

$$y_t'\Gamma = x_t'B \quad + e_t' , \qquad\qquad t = 1, 2, \ldots, n.$$

where Γ is a square, nonsingular matrix. We shall presume that the elements of y_t and x_t are ordered so that the first column of Γ has $L'+1$ leading nonzero entries and the first column of B has k' leading nonzero entries; we shall also presume that $\gamma_{11} = 1$. With these conventions the first equation of the system may be written as

$$y_{1t} = (y_{2t}, y_{3t}, \ldots , y_{L'}, x_{1t}, x_{2t}, \ldots , x_{K'}) \begin{bmatrix} -\gamma_{12} \\ \vdots \\ -\gamma_{1L'} \\ \beta_{11} \\ \vdots \\ \beta_{1k'} \end{bmatrix} + e_{1t}$$

$$= w_t' \, \delta + e_{1t}$$

The linear two-stage least-squares estimator is obtained by putting $z_t = x_t$; that is, the linear two-stage least-squares estimator is the minimizer $\hat{\delta}$ of

$$S(\delta) = [\textstyle\sum_{t=1}^n (y_{1t} - w_t'\delta) \cdot x_t]'(\sum_{t=1}^n x_t x_t')^{-1}[\sum_{t=1}^n (y_{1t} - w_t'\delta) \cdot x_t].$$

Let $\hat{w}_t$ denote the predicted values from a regression of w_t on x_t, viz.

$$\hat{w}_t' = x_t' \left(\sum_{s=1}^{n} x_s x_s'\right)^{-1} \sum_{s=1}^{n} x_s w_s' \qquad t = 1, 2, \ldots, n.$$

Show that a regression of y_{1t} on $\hat{w}_t$ yields $\hat{\delta}$; that is, show that

$$\hat{\delta} = \left(\sum_{t=1}^{n} \hat{w}_t \hat{w}_t'\right)^{-1} \sum_{t=1}^{n} \hat{w}_t y_{1t}.$$

It is from this fact that the name two-stage least squares derives; the first stage is the regression of w_t on x_t and the second is the regression of y_{1t} on $\hat{w}_t$.

3. THE DYNAMIC CASE: GENERALIZED METHOD OF MOMENTS.

Although there is a substantial difference in theory between the dynamic case where errors may be serially correlated and lagged dependent variables may be used as explanatory variables and the regression case where errors are independent and lagged dependent variables are disallowed, there is little difference in applications. All that changes is that the variance V of $n \cdot m_n(\theta^o)$ is estimated differently.

The underlying system is

$$q_\alpha(y_t, x_t, \theta_\alpha^o) = e_{\alpha t} \qquad t = 0, \pm 1, \pm 2, \dots ; \quad \alpha = 1, 2, \dots, L$$

where t indexes observations that are ordered in time, $q_\alpha(y, x, \theta_\alpha)$ is a real-valued function, y_t is an L-vector, x_t is a k-vector, θ_α^o is a p_α-vector of unknown parameters, and $e_{\alpha t}$ is an unobservable observational or experimental error. The vector x_t can include lagged values of the dependent variable $(y_{t-1}, y_{t-2},$ etc.) as elements. Because of these lagged values, x_t is called the vector of predetermined variables rather than the independent variables. The errors $e_{\alpha t}$ will usually be serially correlated

$$\mathcal{C}(e_{\alpha s}, e_{\beta t}) = \sigma_{\alpha\beta st} \neq 0 \qquad \alpha, \beta = 1, 2, \dots, M; \quad s,t = 1, 2, \dots$$

We do not assume that the errors are stationary which accounts for the st index; if the errors were stationary we would have $\sigma_{\alpha\beta}$.

Attention is restricted to the first M equations

$$q_\alpha(y_t, x_t, \theta_\alpha^o) = e_{\alpha t} \qquad t = 1, 2, \dots, n; \quad \alpha = 1, 2, \dots, M.$$

As in the regression case, let θ be a p-vector containing the non-redundant parameters in the set $\theta_1, \theta_2, \dots, \theta_M$ and let

$$q(y,x,\theta) = \begin{bmatrix} q_1(y,x,\theta_1) \\ q_2(y,x,\theta_2) \\ \vdots \\ q_M(y,x,\theta_M) \end{bmatrix}, \qquad e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_M \end{bmatrix},$$

$$q(y_t,x_t,\theta) = \begin{bmatrix} q_1(y_t,x_t,\theta_1) \\ q_2(y_t,x_t,\theta_2) \\ \vdots \\ q_M(y_t,x_t,\theta_M) \end{bmatrix}, \qquad e_t = \begin{bmatrix} e_{1,t} \\ e_{2,t} \\ \vdots \\ e_{M,t} \end{bmatrix}.$$

The analysis unconditional; indeed, the presence of lagged values of y_t as components of x_t precludes a conditional analysis. The theory on which the analysis is based (Chapter 9) does not rely explicitly on the existence of a smooth reduced form as was the case in the previous section. What is required is the existence of measurable functions $W_t(\cdot)$ that depend on the infinite sequence

$$v_\infty = (\ \ldots,\ e_{-1},x_{-1},\ e_0,x_0,\ e_1,x_1,\ \ldots\ )$$

$$= (\ \ldots,\ v_{-1},\ v_0,\ v_1,\ \ldots\ )$$

such that

$$(y_t,x_t) = W_t(v_\infty)\ .$$

and mixing conditions that limit the dependence between (e_s,x_s) and (e_t,x_t) for $t \neq s$. The details are spelled out in Sections 3 and 5 of Chapter 9.

The estimation strategy is the same as nonlinear three-stage least squares (Section 2). One chooses a K-vector of instrumental variables z_t of the form

$$z_t = Z(x_t),$$

forms sample moments

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

where

$$m(y_t, x_t, \theta) = q(y_t, x_t, \theta) \otimes z_t = \begin{bmatrix} q_1(y_t, x_t, \theta_1) \cdot z_t \\ q_2(y_t, x_t, \theta_2) \cdot z_t \\ \vdots \\ q_M(y_t, x_t, \theta_M) \cdot z_t \end{bmatrix}_{M \cdot K \quad 1},$$

and estimates θ^o by that value $\hat{\theta}$ that minimizes

$$S(\theta, V) = [n \cdot m_n(\theta)]' \, V^{-1} [n \cdot m_n(\theta)]$$

with

$$V = \mathcal{C}\left\{[n \cdot m_n(\theta^o)],\ [n \cdot m_n(\theta^o)]'\right\}.$$

In this case, the random variables

$$m(y_t, x_t, \theta^o) = e_t \otimes z_t \qquad t = 1, 2, \ldots, n$$

are correlated and we have

$$V = \mathcal{E}(\Sigma_{t=1}^{n} e_t \otimes z_t)(\Sigma_{s=1}^{n} e_s \otimes z_s)'$$

$$= \Sigma_{t=1}^{n} \Sigma_{s=1}^{n} \mathcal{E}(e_t \otimes z_t)(e_s \otimes z_s)'$$

$$= n \cdot \Sigma_{\tau=-n-1}^{n-1} S_{n\tau}^{o}$$

$$= n \cdot S_n^{o}$$

where

$$S_{n\tau}^{o} = \begin{cases} (1/n)\sum_{t=1+\tau}^{n}\mathcal{E}(e_t \otimes z_t)(e_{t-\tau} \otimes z_{t-\tau})' & \tau \geq 0, \\ (S_{n,-\tau}^{o})' & \tau < 0. \end{cases}$$

To estimate V, we shall need a consistent estimator of S_n^o, that is, an estimator $\hat{S}_n$ with

$$\ell im_{n\to\infty} P(|S_{n,\alpha\beta}^{o} - \hat{S}_{n,\alpha\beta}| > \varepsilon) = 0 \qquad \alpha,\ \beta = 1,\ 2,\ \ldots,\ M$$

for any $\varepsilon > 0$. This is basically a matter of guaranteeing that

$$\ell im_{n\to\infty} \operatorname{Var}(\hat{S}_{n,\alpha\beta}) = 0;$$

see Theorem 3 in Section 2 of Chapter 9.

A consistent estimate $\hat{S}_{n\tau}$ of $S_{n\tau}^{o}$ can be obtained in the obvious way by putting

$$\hat{S}_{n\tau} = \begin{cases} (1/n)\sum_{t=1+\tau}^{n}[q(y_t,x_t,\hat{\theta}^{\#}) \otimes z_t][(q(y_{t-\tau},x_{t-\tau},\hat{\theta}^{\#}) \otimes z_{t-\tau}]' & \tau \geq 0, \\ (\hat{S}_{n,-\tau})' & \tau < 0. \end{cases}$$

where

$$\hat{\theta}^{\#} = \operatorname{argmin}_{\theta} S(\theta,\ I \otimes \textstyle\sum_{t=1}^{n} z_t z_t').$$

However, one cannot simply add the $\hat{S}_{n\tau}$ for τ ranging from $-(n-1)$ to $(n-1)$ as suggested by the definition of S_n^o and obtain a consistent estimator because $\operatorname{Var}(\hat{S}_n)$ will not decrease with n. The variance will decrease if a smaller number of summands is used, namely the sum for τ ranging from $-\ell(n)$ to $\ell(n)$ where $\ell(n)$ = the integer nearest $n^{1/5}$.

Consistency will obtain with this modification but the unweighted sum will not be positive definite in general. As we propose to minimize $S(\theta,\hat{V})$ in order to compute $\hat{\theta}$, the matrix $\hat{V} = n \cdot \hat{S}_n$ must be positive definite. The weighted sum

$$\hat{S}_n = \sum_{\tau=-\ell(n)}^{\ell(n)} w[\tau/\ell(n)]\ \hat{S}_{n\tau}$$

constructed from Parzen weights

$$w(x) = \begin{cases} 1 - 6|x|^2 + 6|x|^3 & 0 \le x \le 1/2 \\ 2(1 - |x|)^3 & 1/2 \le x \le 1. \end{cases}$$

is consistent and positive definite (Theorem 3, Section 2, Chapter 3). The motivation for this particular choice of weights derives from the observation that if $\{e_t \otimes z_t\}$ were a stationary time series then S_n^o would be the spectral density of the process evaluated at zero; the estimator with Parzen weights is the best estimator of the spectral density in an asymptotic mean square error sense (Anderson, 1971, Chapter 9 or Bloomfield, 1976, Chapter 7).

The generalized method of moments differs from the three-stage least-squares estimator only in the computation of $\hat{V}$. The rest is the same, the estimate of θ^o is the minimizer $\hat{\theta}$ of $S(\theta,\hat{V})$, viz.

$$\hat{\theta} = \operatorname{argmin}_{\theta}\ S(\theta,\hat{V}).$$

the estimated variance-covariance of $\hat{\theta}$ is

$$\hat{C} = \left\{[\sum_{t=1}^{n} Q(y_t,x_t,\hat{\theta}) \otimes z_t]'\hat{V}^{-1}[\sum_{t=1}^{n} Q(y_t,x_t,\hat{\theta}) \otimes z_t]\right\}^{-1}$$

where

$$Q(y_t,x_t,\theta) = (\partial/\partial\theta')q(y_t,x_t,\theta) ,$$

and, the Gauss-Newton correction vector is

$$D(\theta,V) = -\left\{[\textstyle\sum_{t=1}^{n} Q(y_t,x_t,\theta) \otimes z_t]'V^{-1}[\sum_{t=1}^{n} Q(y_t,x_t,\theta) \otimes z_t]\right\}^{-1}$$
$$\times \left\{[\textstyle\sum_{t=1}^{n} Q(y_t,x_t,\theta) \otimes z_t]'V^{-1}[\sum_{t=1}^{n} q(y_t,x_t,\theta) \otimes z_t]\right\} .$$

We illustrate.

EXAMPLE 2 (Continued). Recall that

$$q(y_t,x_t,\theta) = \beta(y_t)^{\alpha} x_t - 1 \qquad t = 1, 2, \ldots, 239.$$

where, taking values from Table 1a,

$$y_t = \frac{\text{consumption at time t / population at time t}}{\text{consumption at time t - 1 / population at time t - 1}}$$

$$x_t = (1 + \text{stock returns at time t}) \frac{\text{deflator at time t - 1}}{\text{deflator at time t}} .$$

The instrumental variables employed in the estimation are

$$z_t = (1, y_{t-1}, x_{t-1})'.$$

or instance, we will have

$$y_2 = 1.01061, \quad x_2 = 1.00628, \quad \text{and} \quad z_2 = (1, 1.00346, 1.008677)'.$$

Recall also that, in theory,

$$\mathcal{E}\,(e_t \otimes z_t) = 0, \qquad t = 2, 3, \ldots, 239,$$

$$\mathcal{E}\,(e_t \otimes z_t)(e_t \otimes z_t)' = \Sigma, \qquad t = 2, 3, \ldots, 239,$$

$$\mathcal{E}\,(e_t \otimes z_t)(e_t \otimes z_t)' = 0, \qquad t \neq s.$$

Because the variance estimator has the form

$$\hat{V} = n \cdot \sum_{\tau=-\ell(n)}^{\ell(n)} w[\tau/\ell(n)]\, \hat{S}_{n\tau}$$

where

$$\ell(n) = (n)^{1/5},$$

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

and

$$\hat{S}_{n\tau} = \begin{cases} (1/n) \sum_{t=1+\tau}^{n} [q(y_t, x_t, \hat{\theta}^{\#}) \otimes z_t][(q(y_{t-\tau}, x_{t-\tau}, \hat{\theta}^{\#}) \otimes z_{t-\tau}]' & \tau \geq 0, \\ (\hat{S}_{n,-\tau})' & \tau < 0. \end{cases}$$

whereas PROC SYSNLIN can only compute a variance estimate of the form

$$\hat{V} = \left[(1/n)\sum_{t=1}^{n} q(y_t, x_t, \hat{\theta}^{\#})\; q'(y_t, x_t, \hat{\theta}^{\#})\right] \otimes \left[\sum_{t=1}^{n} z_t z_t'\right]$$

we are sort of on our own as far as writing code is concerned. Our strategy will be to use PROC MATRIX using

```
DATA WORK01;  SET EG02;
NDSPER=NDS/PEOPLE; Y=NDSPER/LAG(NDSPER); X=(1+STOCKS)*LAG(DEFLATOR)/DEFLATOR;
DATA WORK02;  SET WORK01;  Z0=1; Z1=LAG(Y); Z2=LAG(X);  IF _N_=1 THEN DELETE;
PROC MATRIX;  FETCH Y DATA=WORK02(KEEP=Y);  FETCH X DATA=WORK02(KEEP=X);
              FETCH Z DATA=WORK02(KEEP=Z0 Z1 Z2); Z(1,)=0 0 0;
A=-.4; B=.9; V=Z'*Z; %GAUSS DO WHILE (S>OBJ#(1+1.E-5)); %GAUSS END;
TSHARP=A // B;  PRINT TSHARP;
%VARIANCE  V=VHAT;  %GAUSS DO WHILE (S>OBJ#(1+1.E-5)); %GAUSS END;
THAT=A // B;  PRINT VHAT THAT CHAT S;
```

where %GAUSS is a MACRO which computes a modified (line searched) Gauss-Newton iterative step,

```
%MACRO GAUSS;
 M=0/0/0; DELM=0 0/0 0/0 0; ONE=1;
 DO T=2 TO 239;
  QT = B#Y(T,1)##A#X(T,1)-ONE;
  DELQTA = B#LOG(Y(T,1))#Y(T,1)##A#X(T,1);  DELQTB = Y(T,1)##A#X(T,1);
  MT = QT @ Z(T,)%STR(%');  DELMT = (DELQTA || DELQTB) @ Z(T,)%STR(%');
  M=M+MT;  DELM=DELM+DELMT;
 END;
 CHAT=INV(DELM%STR(%')*INV(V)*DELM);  D=-CHAT*DELM%STR(%')*INV(V)*M;
 S=M%STR(%')*INV(V)*M;  OBJ=S;  L=2;  COUNT=0;  A0=A;  B0=B;
 DO WHILE (OBJ>=S & COUNT<=40);
  COUNT=COUNT+ONE;  L=L#.5;  A=A0+L#D(1,1);  B=B0+L#D(2,1);  M=0;
  DO T=2 TO 239;  M=M+(B#Y(T,1)##A#X(T,1)-ONE)@Z(T,)%STR(%');  END;
  OBJ=M%STR(%')*INV(V)*M;
 END;
%MEND GAUSS;
```

and %VARIANCE is a MACRO which computes a variance estimate

```
%MACRO VARIANCE;
 S0=0 0 0/0 0 0/0 0 0;  S1=0 0 0/0 0 0/0 0 0; S2=0 0 0/0 0 0/0 0 0; ONE=1;
 DO T=2 TO 239;
  MT0=(B#Y(T,1)##A#X(T,1)-ONE) @ Z(T,)%STR(%'); S0=S0+MT0*MT0%STR(%');
  IF T>3 THEN DO;
  MT1=(B#Y(T-1,1)##A#X(T-1,1)-ONE)@Z(T-1,)%STR(%'); S1=S1+MT0*MT1%STR(%'); END;
  IF T>4 THEN DO;
  MT2=(B#Y(T-2,1)##A#X(T-2,1)-ONE)@Z(T-2,)%STR(%'); S2=S2+MT0*MT2%STR(%'); END;
 END;
 W0=1;  W1=ONE-6#(1#/3)##2+6#(1#/3)##3;  W2=2#(ONE-(2#/3))##3;  W3=0;
 VHAT=(W0#S0+W1#S1+W1#S1%STR(%')+W2#S2+W2#S2%STR(%'));
%MEND VARIANCE;
```

The code is fairly transparent if one takes a one by one matrix to be a scalar and reads ' for %STR(%'), * for #, ** for ##, / for #/, and $\otimes$ for @.

While in general this code is correct, for this particular problem

$$\mathcal{E}\,(e_t \otimes z_t)(e_t \otimes z_t)' = 0, \qquad t \neq s.$$

so we shall replace the line

```
WO=1;  W1=ONE-6#(1#/3)##2+6#(1#/3)##3;  W2=2#(ONE-(2#/3))##3;  W3=0;
```

in %VARIANCE which computes the weights $w(\tau/\ell(n))$ with the line

```
WO=1;  W1=0;  W2=0;  W3=0;
```

The computations are shown in Figure 2.

Figure 2. The Generalized Method of Moments Estimator for Example 2.

SAS Statements:

```
%MACRO GAUSS;
 M=0/0/0; DELM=0 0/0 0/0 0; ONE=1;
 DO T=2 TO 239;
  QT = B#Y(T,1)##A#X(T,1)-ONE;
  DELQTA = B#LOG(Y(T,1))#Y(T,1)##A#X(T,1);  DELQTB = Y(T,1)##A#X(T,1);
  MT = QT @ Z(T,)%STR(%');  DELMT = (DELQTA || DELQTB) @ Z(T,)%STR(%');
  M=M+MT;  DELM=DELM+DELMT;
 END;
 CHAT=INV(DELM%STR(%')*INV(V)*DELM);  D=-CHAT*DELM%STR(%')*INV(V)*M;
 S=M%STR(%')*INV(V)*M;  OBJ=S;  L=2;  COUNT=0;  A0=A;  B0=B;
 DO WHILE (OBJ>=S & COUNT<=40);
  COUNT=COUNT+ONE;  L=L#.5;  A=A0+L#D(1,1);  B=B0+L#D(2,1);  M=0;
  DO T=2 TO 239;  M=M+(B#Y(T,1)##A#X(T,1)-ONE)@Z(T,)%STR(%');  END;
  OBJ=M%STR(%')*INV(V)*M;
 END;
%MEND GAUSS;
%MACRO VARIANCE;
 S0=0 0 0/0 0 0/0 0 0;  S1=0 0 0/0 0 0/0 0 0; S2=0 0 0/0 0 0/0 0 0; ONE=1;
 DO T=2 TO 239;
  MT0=(B#Y(T,1)##A#X(T,1)-ONE) @ Z(T,)%STR(%'); S0=S0+MT0*MT0%STR(%');
  IF T>3 THEN DO;
  MT1=(B#Y(T-1,1)##A#X(T-1,1)-ONE)@Z(T-1,)%STR(%'); S1=S1+MT0*MT1%STR(%'); END;
  IF T>4 THEN DO;
  MT2=(B#Y(T-2,1)##A#X(T-2,1)-ONE)@Z(T-2,)%STR(%'); S2=S2+MT0*MT2%STR(%'); END;
 END;
 W0=1;  W1=0;  W2=0;  W3=0;
 VHAT=(W0#S0+W1#S1+W1#S1%STR(%')+W2#S2+W2#S2%STR(%'));
%MEND VARIANCE;
DATA WORK01;  SET EG02;
NDSPER=NDS/PEOPLE; Y=NDSPER/LAG(NDSPER); X=(1+STOCKS)*LAG(DEFLATOR)/DEFLATOR;
DATA WORK02;  SET WORK01;  Z0=1; Z1=LAG(Y); Z2=LAG(X);  IF _N_=1 THEN DELETE;
PROC MATRIX;  FETCH Y DATA=WORK02(KEEP=Y);  FETCH X DATA=WORK02(KEEP=X);
              FETCH Z DATA=WORK02(KEEP=Z0 Z1 Z2); Z(1,)=0 0 0;
A=-.4; B=.9; V=Z'*Z; %GAUSS DO WHILE (S>OBJ#(1+1.E-5)); %GAUSS END;
TSHARP=A // B;  PRINT TSHARP;
%VARIANCE  V=VHAT;  %GAUSS DO WHILE (S>OBJ#(1+1.E-5)); %GAUSS END;
THAT=A // B;  PRINT VHAT THAT CHAT;
```

Figure 2. (Continued).

Output:

SAS 1

TSHARP	COL1
ROW1	-0.848852
ROW2	0.998929

VHAT	COL1	COL2	COL3
ROW1	0.405822	0.406434	0.398737
ROW2	0.406434	0.407055	0.399363
ROW3	0.398737	0.399363	0.392723

THAT	COL1
ROW1	-1.03352
ROW2	0.998256

CHAT	COL1	COL2
ROW1	3.58009	-0.00721267
ROW2	-0.00721267	.0000206032

S	COL1
ROW1	1.05692

Tauchen (1986) considers the question of how instruments ought to be chosen for generalized method of moments estimators in the case where the errors are uncorrelated. In this case the optimal choice of instrumental variables is (Hansen, 1985)

$$z_t = \mathcal{E}_t(\partial/\partial\theta')q(y_t,x_t,\theta^o)[\mathcal{E}_t(e_t e_t')]^{-1}$$

where $\mathcal{E}_t(\cdot)$ denotes the conditional expectation with respect to all variables (information) relevant to the problem from the present time t to as far into the past as is relevant; see the discussion of this point in Example 2 of Section 2. Tauchen, using the same sort of model as Example 2, obtains the small sample bias and variance for various choices of instrumental variables which he compares to the optimal choice. He finds that, when short lag lengths are used in forming instrumental variables, nearly asymptotically optimal parameter estimates obtain and that, as lag length increases, estimates become increasingly concentrated around biased values and confidence intervals become increasingly inaccurate. He also finds that the test of overidentifying restrictions performs reasonably well in finite samples.

The more interesting aspect of Tauchen's work is that he obtains a computational strategy for generating data that follows a nonlinear, dynamic model that can be used to formulate a bootstrap strategy to find the optimal instrumental variables in a given application.

PROBLEMS

1. Use the data of Tables 1a and 1b of Section 1 to reproduce the results of Hansen and Singleton (1984).

2. Verify that if one uses the first order conditions for three month treasury bills, $z_t = (1, y_{t-s}, x_{t-s})$ with s chosen the smallest value that will ensure that $\mathcal{E}(q(y_t, x_t, \theta^o) \otimes z_t) = 0$, and Parzen weights then

TSHARP	COL1
ROW1	-4.38322
ROW2	1.02499

VHAT	COL1	COL2	COL3
ROW1	0.24656	0.248259	0.246906
ROW2	0.248259	0.249976	0.248609
ROW3	0.246906	0.248609	0.247256

THAT	COL1
ROW1	-4.37803
ROW2	1.02505

CHAT	COL1	COL2
ROW1	22.8898	-0.140163
ROW2	-0.140163	0.00086282

Should Parzen weights be used in this instance?

4. HYPOTHESIS TESTING

As seen in the last two sections, the effect of various assumptions regarding lagged dependent variables, heteroskedasticity, or autocorrelated errors is to alter the form of the variance estimator $\hat{V}$ without affecting the form of the estimator $\hat{\theta}$. Thus, each estimator can be regarded as, at most, a simplified version of the general estimator proposed in Section 5 of Chapter 9 and, in consequence, the theory of hypothesis testing presented in Section 6 of Chapter 9 applies to all of them. This being the case, here we can lump the preceding estimation procedures together and accept the following as the generic description of the hypothesis testing problem.

Attention is restricted to the first M equations

$$q_{\alpha t}(y_t, x_t, \theta_\alpha^o) = e_{\alpha t} \qquad t = 1, 2, \ldots, n; \quad \alpha = 1, 2, \ldots, M.$$

of some system. Let θ be a p-vector containing the non-redundant parameters in the set $\theta_1, \theta_2, \ldots, \theta_M$ and let

$$q(y,x,\theta) = \begin{bmatrix} q_1(y,x,\theta_1) \\ q_2(y,x,\theta_2) \\ \vdots \\ q_M(y,x,\theta_M) \end{bmatrix}, \qquad e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_M \end{bmatrix}.$$

To estimate θ^o, one chooses a K-vector of instrumental variables z_t of the form

$$z_t = Z(x_t),$$

constructs the sample moments

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

with

$$m(y_t,x_t,\theta) = q(y_t,x_t,\theta) \otimes z_t = \begin{bmatrix} q_1(y_t,x_t,\theta_1)\cdot z_t \\ q_2(y_t,x_t,\theta_2)\cdot z_t \\ \vdots \\ q_M(y_t,x_t,\theta_M)\cdot z_t \end{bmatrix}_{M\cdot K \; 1},$$

and estimates θ^o by the value $\hat{\theta}$ that minimizes

$$S(\theta,\hat{V}) = [n\cdot m_n(\theta)]' \, \hat{V}^{-1}[n\cdot m_n(\theta)]$$

where $\hat{V}$ is some consistent estimate of

$$V = \mathcal{C}\left\{[n\cdot m_n(\theta^o)],\ [n\cdot m_n(\theta^o)]'\right\} .$$

The estimated variance-covariance of $\hat{\theta}$ is

$$\hat{C} = \left\{[\textstyle\sum_{t=1}^{n} Q(y_t,x_t,\hat{\theta}) \otimes z_t]'\hat{V}^{-1}[\textstyle\sum_{t=1}^{n} Q(y_t,x_t,\hat{\theta}) \otimes z_t]\right\}^{-1}$$

where

$$Q(y_t,x_t,\theta) = (\partial/\partial\theta')q(y_t,x_t,\theta) .$$

The Gauss-Newton correction vector is

$$D(\theta,V) = -\left\{[\textstyle\sum_{t=1}^{n} Q(y_t,x_t,\theta) \otimes z_t]'V^{-1}[\textstyle\sum_{t=1}^{n} Q(y_t,x_t,\theta) \otimes z_t]\right\}^{-1}$$
$$\times \left\{[\textstyle\sum_{t=1}^{n} Q(y_t,x_t,\theta) \otimes z_t]'V^{-1}[\textstyle\sum_{t=1}^{n} q(y_t,x_t,\theta) \otimes z_t]\right\} .$$

With this as a backdrop, interest centers in testing a hypothesis that can be expressed either as a parametric restriction

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

or as a functional dependency

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

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

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

which is assumed to be continuous with rank q at θ^o; $g(\rho)$ 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

$$q(y,x,\theta) = \begin{bmatrix} y_1 - \ell n \dfrac{\theta_1 + \theta_2 \cdot r_1 + \theta_3 \cdot r_2 + \theta_4 \cdot r_3 - (\theta_2 + \theta_3 + \theta_4) \cdot y_3}{-1 + \theta_4 \cdot r_1 + \theta_7 \cdot r_2 + \theta_8 \cdot r_3 - (\theta_4 + \theta_7 + \theta_8) \cdot y_3} \\[2ex] y_2 - \ell n \dfrac{\theta_5 + \theta_3 \cdot r_1 + \theta_6 \cdot r_2 + \theta_7 \cdot r_3 - (\theta_3 + \theta_6 + \theta_7) \cdot y_3}{-1 + \theta_4 \cdot r_1 + \theta_7 \cdot r_2 + \theta_8 \cdot r_3 - (\theta_4 + \theta_7 + \theta_8) \cdot y_3} \end{bmatrix}$$

with

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

The hypothesis of homogeneity, see Section 4 of Chapter 6, may be written as the parametric restriction

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

with Jacobian

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

or, equivalently, as the functional dependency

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

with Jacobian

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

The Wald test statistic for the hypothesis

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

is

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

where $\hat{h} = h(\hat{\theta})$, $H(\theta) = (\partial/\partial\theta')\,h(\theta)$, and $\hat{H} = H(\hat{\theta})$. One rejects the hypothesis

$$H\colon h(\theta^{o}) = 0$$

when W exceeds the upper $\alpha \times 100\%$ critical point χ^2_α of the chi-square distribution with q degrees of freedom; $\chi^2_\alpha = (\chi^2)^{-1}(1-\alpha,\ q)$.

Under the alternative $A\colon h(\theta^{o}) \neq 0$, the Wald test statistic is approximately distributed as the non-central chi-square with q degrees of freedom and non-centrality parameter

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

where

$$C = \left\{\sum_{t=1}^{n}\mathcal{E}[Q(y_t,x_t,\theta^{o}) \otimes Z(x_t)]'V^{-1}\sum_{t=1}^{n}\mathcal{E}[Q(y_t,x_t,\theta^{o}) \otimes Z(x_t)]\right\}^{-1},$$

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

$$V = \mathcal{C}\left\{[n\cdot m_n(\theta^{o})],\ [n\cdot m_n(\theta^{o})]'\right\}\ ,$$

$$Q(y_t,x_t,\theta) = (\partial/\partial\theta')q(y_t,x_t,\theta)\ .$$

Note in the formulas above, that if x_t is random then the expectation is

$\mathcal{E}[Q(y_t,x_t,\theta^o) \otimes Z(x_t)]$, not $\mathcal{E}[Q(y_t,x_t,\theta^o)] \otimes Z(x_t)$. If there are no lagged dependent variables (and the analysis is conditional), these two expectations will be the same.

EXAMPLE 1. (continued). Code to compute the Wald test statistic

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

for the hypothesis of homogeneity

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

is shown in Figure 3. The nonlinear three-stage least-squares estimators $\hat{\theta}$ and $\hat{C}$ are computed using the same code as in Figure 1 of Section 3. The computed values are passed to PROC MATRIX where the value

$$W = 3.01278 \qquad \text{(from Figure 3)}$$

is computed using straightforward algebra. Since $(\chi^2)^{-1}(.95,3) = 7.815$, the hypothesis is accepted at the 5% level.

In Section 4 of Chapter 6, using the multivariate least-squares estimator, the hypothesis was rejected. The conflicting results are due to the larger estimated variance with which the three-stage least-squares estimator is computed with these data. As remarked earlier, the multivariate least squares estimator is computed from higher quality data than the three-stage least-squares estimator in this instance so that the multivariate least-squares results are more credible.∎

Figure 3. Illustration of Wald Test Computations with Example 1.

SAS Statements:

```
PROC MODEL OUT=MOD01;
ENDOGENOUS Y1 Y2 Y3;
EXOGENOUS  R1 R2 R3 D0 D1 D2 D3 D4 D5 D6 D7 D8 D9 D10 D11 D12 D13;
PARMS T1 T2 T3 T4 T5 T6 T7 T8 C0 C1 C2 C3 C4 C5 C6 C7 C8 C9 C10 C11 C12 C13;
PEAK= T1+T2*R1+T3*R2+T4*R3-(T2+T3+T4)*Y3;
INTER=T5+T3*R1+T6*R2+T7*R3-(T3+T6+T7)*Y3;
BASE= -1+T4*R1+T7*R2+T8*R3-(T4+T7+T8)*Y3;
Y1=LOG(PEAK/BASE);  Y2=LOG(INTER/BASE);
Y3=D0*C0+D1*C1+D2*C2+D3*C3+D4*C4+D5*C5+D6*C6+D7*C7+D8*C8+D9*C9+D10*C10+D11*C11
   +D12*C12+D13*C13;

PROC SYSNLIN DATA=EG01 MODEL=MOD01 N3SLS METHOD=GAUSS MAXIT=50 CONVERGE=1.E-8
             SDATA=IDENTITY OUTS=SHAT OUTEST=TSHARP;
INSTRUMENTS  R1 R2 R3 D0 D1 D2 D3 D4 D5 D6 D7 D8 D9 D10 D11 D12 D13 / NOINT;
FIT Y1 Y2 START = (T1 -2.98  T2 -1.16  T3 0.787  T4 0.353  T5 -1.51  T6 -1.00
                   T7 0.054  T8 -0.474);
PROC SYSNLIN DATA=EG01 MODEL=MOD01 N3SLS METHOD=GAUSS MAXIT=50 CONVERGE=1.E-8
             SDATA=SHAT ESTDATA=TSHARP OUTEST=WORK01 COVOUT;
INSTRUMENTS  R1 R2 R3 D0 D1 D2 D3 D4 D5 D6 D7 D8 D9 D10 D11 D12 D13 / NOINT;
FIT Y1 Y2;

PROC MATRIX; FETCH W DATA=WORK01(KEEP = T1 T2 T3 T4 T5 T6 T7 T8);
THAT=W(1,)';  CHAT=W(2:9,);
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;
W=THAT'*H'*INV(H*CHAT*H')*H*THAT;  PRINT W;
```

Output:

```
                              SAS                                    8


               W                 COL1

               ROW1           3.01278
```

Let $\tilde{\theta}$ denote the value of θ that minimizes $S(\theta,\hat{V})$ subject to $h(\theta) = 0$. Equivalently, let $\hat{\rho}$ denote the value of ρ that achieves the unconstrained minimum of $S(g(\rho),\hat{V})$ and put $\tilde{\theta} = g(\hat{\rho})$. The "likelihood ratio" test statistic for the hypothesis

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

is

$$L = S(\tilde{\theta},\hat{V}) - S(\hat{\theta},\hat{V}).$$

It is essential that $\hat{V}$ be the same matrix in both terms on the right hand side, they must be exactly the same not just "asymptotically equivalent."

One rejects $H\colon h(\theta^{o}) = 0$ when L exceeds the upper $\alpha \times 100\%$ critical point χ^2_α of the chi-square distribution with q degrees of freedom; $\chi^2_\alpha = (\chi^2)^{-1}(1-\alpha,\ q)$.

Let θ^* denote the value of θ that minimizes

$$S^{o}(\theta,V) = [n\cdot\mathcal{E}m_n(\theta)]'\ V^{-1}[n\cdot\mathcal{E}m_n(\theta)]$$

subject to

$$h(\theta) = 0.$$

Equivalently, let ρ^{o} denote the value of ρ that achieves the unconstrained minimum of $S^{o}(g(\rho),V)$ and put $\theta^* = g(\rho^{o})$.

Under the alternative, $A\colon h(\theta^{o}) \neq 0$, the "likelihood ratio" test statistic L is approximately distributed as the non-central chi-square with q degrees of freedom and non-centrality parameter

$$\lambda = \left\{\textstyle\sum_{t=1}^{n} \mathcal{E}[q(y_t, x_t, \theta^*) \otimes Z(x_t)]\right\}' V^{-1} \left\{\textstyle\sum_{t=1}^{n} \mathcal{E}[Q(y_t, x_t, \theta^*) \otimes Z(x_t)]\right\}$$

$$\times\ J^{-1} H' (H J^{-1} H') H J^{-1}$$

$$\times \left\{\textstyle\sum_{t=1}^{n} \mathcal{E}[Q(y_t, x_t, \theta^*) \otimes Z(x_t)]\right\}' V^{-1} \left\{\textstyle\sum_{t=1}^{n} \mathcal{E}[q(y_t, x_t, \theta^*) \otimes Z(x_t)]\right\} / 2$$

where

$$V = \mathcal{C}\left\{[n \cdot m_n(\theta^o)],\ [n \cdot m_n(\theta^o)]'\right\} ,$$

$$Q(y_t, x_t, \theta) = (\partial/\partial\theta') q(y_t, x_t, \theta) ,$$

$$J = \left\{\textstyle\sum_{t=1}^{n} \mathcal{E}[Q(y_t, x_t, \theta^*) \otimes Z(x_t)]\right\}' V^{-1} \left\{\textstyle\sum_{t=1}^{n} \mathcal{E}[Q(y_t, x_t, \theta^*) \otimes Z(x_t)]\right\} ,$$

$$H = H(\theta^*) = (\partial/\partial\theta')\ h(\theta^*) .$$

Alternative expressions for λ can be obtained using Taylor's theorem and the relationship $H'(HJ^{-1}H')^{-1}H = J - JG(G'JG)^{-1}G'J$ from Section 6 of Chapter 3; see Gallant and Jorgenson (1979).

EXAMPLE 1. (continued). The hypothesis of homogeneity in the model

$$q(y, x, \theta) = \begin{bmatrix} y_1 - \ln \dfrac{\theta_1 + \theta_2 \cdot r_1 + \theta_3 \cdot r_2 + \theta_4 \cdot r_3 - (\theta_2 + \theta_3 + \theta_4) \cdot y_3}{-1 + \theta_4 \cdot r_1 + \theta_7 \cdot r_2 + \theta_8 \cdot r_3 - (\theta_4 + \theta_7 + \theta_8) \cdot y_3} \\ \\ y_2 - \ln \dfrac{\theta_5 + \theta_3 \cdot r_1 + \theta_6 \cdot r_2 + \theta_7 \cdot r_3 - (\theta_3 + \theta_6 + \theta_7) \cdot y_3}{-1 + \theta_4 \cdot r_1 + \theta_7 \cdot r_2 + \theta_8 \cdot r_3 - (\theta_4 + \theta_7 + \theta_8) \cdot y_3} \end{bmatrix}$$

can be expressed as the functional dependency

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

Minimization of $S[g(\rho),\hat{V}]$ as shown in Figure 4 gives

$$S(\tilde{\theta},\hat{V}) = 38.34820 \qquad \text{(from Figure 4)}$$

and we have

$$S(\hat{\theta},\hat{V}) = 34.96403 \qquad \text{(from Figure 1)}$$

from Figure 1 of Section 2. Thus

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

$$= 38.34820 - 34.96403$$

$$= 3.38417.$$

Since $(\chi^2)^{-1}(.95,3) = 7.815$, the hypothesis is accepted at the 5% level. Note in Figures 1 and 4 that $\hat{V}$ is computed the same. As mentioned several times, the test is invalid if care is not taken to be certain that this is so. ▮

Figure 4. Example 1 Fitted by Nonlinear Three-Stage Least Squares, Homogeneity Imposed.

SAS Statements:

```
PROC MODEL OUT=MOD02;
ENDOGENOUS Y1 Y2 Y3;
EXOGENOUS  R1 R2 R3 D0 D1 D2 D3 D4 D5 D6 D7 D8 D9 D10 D11 D12 D13;
PARMS RO1 RO2 RO3 RO4 RO5 C0 C1 C2 C3 C4 C5 C6 C7 C8 C9 C10 C11 C12 C13;
T1=RO1; T2=-RO2-RO3; T3=RO2; T4=RO3; T5=RO4; T6=-RO5-RO2; T7=RO5; T8=-RO5-RO3;
PEAK= T1+T2*R1+T3*R2+T4*R3-(T2+T3+T4)*Y3;
INTER=T5+T3*R1+T6*R2+T7*R3-(T3+T6+T7)*Y3;
BASE= -1+T4*R1+T7*R2+T8*R3-(T4+T7+T8)*Y3;
Y1=LOG(PEAK/BASE);  Y2=LOG(INTER/BASE);
Y3=D0*C0+D1*C1+D2*C2+D3*C3+D4*C4+D5*C5+D6*C6+D7*C7+D8*C8+D9*C9+D10*C10+D11*C11
   +D12*C12+D13*C13;
PROC SYSNLIN DATA=EG01 MODEL=MOD02 N3SLS METHOD=GAUSS MAXIT=50 CONVERGE=1.E-8
             SDATA=SHAT;
INSTRUMENTS  R1 R2 R3 D0 D1 D2 D3 D4 D5 D6 D7 D8 D9 D10 D11 D12 D13 / NOINT;
FIT Y1 Y2 START = (RO1 -3 RO2 .8 RO3 .4 RO4 -1.5 RO5 .03);
```

Output:

```
                                  SAS                                         8

                     NONLINEAR 3SLS PARAMETER ESTIMATES

                                  APPROX.         'T'    APPROX.
           PARAMETER    ESTIMATE  STD ERROR      RATIO   PROB>!T!

           RO1          -2.66573    0.17608     -15.14    0.0001
           RO2           0.84953    0.06641      12.79    0.0001
           RO3           0.37591    0.02686      13.99    0.0001
           RO4          -1.56635    0.07770     -20.16    0.0001
           RO5           0.06129    0.03408       1.80    0.0735


           NUMBER OF OBSERVATIONS        STATISTICS  FOR SYSTEM
           USED                220       OBJECTIVE       0.17431
           MISSING               4       OBJECTIVE*N    38.34820


            COVARIANCE OF RESIDUALS MATRIX USED FOR ESTIMATION

                      S               Y1              Y2

                      Y1         0.17159         0.09675
                      Y2         0.09675         0.09545
```

The Lagrange multiplier test is most apt to be used when the constrained estimator $\tilde{\theta}$ is much easier to compute than the unconstrained estimator $\hat{\theta}$ so it is somewhat unreasonable to expect that a variance estimate $\hat{V}$ computed from unconstrained residuals will be available. Accordingly, let

$$\tilde{\theta}^{\#} = \operatorname{argmin}_{h(\theta)=0} s(\theta, I \otimes \textstyle\sum_{t=1}^{n} z_t z_t').$$

If the model is a pure regression situation put

$$\tilde{V} = \left[(1/n)\textstyle\sum_{t=1}^{n} q(y_t, x_t, \tilde{\theta}^{\#})\, q'(y_t, x_t, \tilde{\theta}^{\#})\right] \otimes \left[\textstyle\sum_{t=1}^{n} z_t z_t'\right] ;$$

if model is the regression situation with heteroskedastic errors put

$$\tilde{V} = \textstyle\sum_{t=1}^{n}\left[q(y_t,x_t,\tilde{\theta}^{\#}) \otimes z_t\right]\left[q(y_t,x_t,\tilde{\theta}^{\#}) \otimes z_t\right]' ;$$

or if the model is dynamic put

$$\tilde{V} = n\bullet\tilde{S}_n$$

where

$$\tilde{S}_n = \textstyle\sum_{\tau=-\ell(n)}^{\ell(n)} w[\tau/\ell(n)]\, \tilde{S}_{n\tau}$$

$$w(x) = \begin{cases} 1 - 6|x|^2 + 6|x|^3 & 0 \le x \le 1/2 \\ 2(1 - |x|)^3 & 1/2 \le x \le 1. \end{cases}$$

$$\tilde{S}_{n\tau} = \begin{cases} (1/n)\ \sum_{t=1+\tau}^{n}[q(y_t,x_t,\tilde{\theta}^{\#}) \otimes z_t][(q(y_{t-\tau},x_{t-\tau},\tilde{\theta}^{\#}) \otimes z_{t-\tau}]' & \tau \ge 0, \\ (\tilde{S}_{n,-\tau})' & \tau < 0. \end{cases}$$

Let

$$\tilde{\tilde{\theta}} = \operatorname{argmin}_{h(\theta) = 0} S(\theta, \tilde{\tilde{V}}).$$

The Gauss-Newton step away from $\tilde{\tilde{\theta}}$ (presumably) toward $\hat{\theta}$ is

$$\tilde{D} = D(\tilde{\tilde{\theta}}, \tilde{V})$$

The Lagrange multiplier test statistic for the hypothesis

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

is

$$R = \tilde{D}'\tilde{J}^{-1}\tilde{D}$$

where

$$\tilde{J} = [\Sigma_{t=1}^{n} Q(y_t, x_t, \tilde{\tilde{\theta}}) \otimes z_t]'\tilde{V}^{-1}[\Sigma_{t=1}^{n} Q(y_t, x_t, \tilde{\tilde{\theta}}) \otimes z_t] \ ,$$

One rejects $H\colon h(\theta^{o}) = 0$ when R exceeds the upper $\alpha \times 100\%$ critical point χ^2_α of the Chi-square distribution with q degrees of freedom; $\chi^2_\alpha = (\chi^2)^{-1}(1-\alpha,\ q)$.

The approximate non-null distribution of the Lagrange multiplier test statistic is the same as the non-null distribution of the "likelihood ratio" test statistic.

EXAMPLE 1. (continued). The computations for the Lagrange multiplier test of homogeneity are shown in Figure 5. As in Figure 4, PROC MODEL defines the model $q[(y,x,g(\rho)]$. In Figure 5, the first use of PROC SYSNLIN computes

$$\hat{\rho}^{\#} = \text{argmin}_{\rho}\ S(\theta,\ I \otimes \textstyle\sum_{t=1}^{n} z_t z_t').$$

$$\tilde{V} = \left[(1/n)\textstyle\sum_{t=1}^{n} q(y_t,\ x_t,\ \tilde{\theta}^{\#})\ q'(y_t,\ x_t,\ \tilde{\theta}^{\#})\right] \otimes \left[\textstyle\sum_{t=1}^{n} z_t z_t'\right]$$

where

$$\tilde{\theta}^{\#} = g(\hat{\rho}^{\#}).$$

The second use of PROC SYSNLIN computes

$$\tilde{\hat{\rho}} = \text{argmin}_{\rho}\ S[g(\rho),\tilde{V}].$$

The subsequent DATA W01 statement computes

$$\tilde{q}_t = q(y_t,x_t,\tilde{\tilde{\theta}})$$

$$\tilde{Q}_t = (\partial/\partial\theta')q(y_t,x_t,\tilde{\tilde{\theta}})$$

where

$$\tilde{\tilde{\theta}} = g(\hat{\rho}^{\#}).$$

Finally PROC MATRIX is used to compute

$$\tilde{J} = [\textstyle\sum_{t=1}^{n} \tilde{Q}_t \otimes z_t]'\tilde{V}^{-1}[\textstyle\sum_{t=1}^{n} \tilde{Q}_t \otimes z_t]\ ,$$

$$\tilde{D} = \tilde{J}^{-1}[\textstyle\sum_{t=1}^{n} \tilde{Q}_t \otimes z_t]'\tilde{V}^{-1}[\textstyle\sum_{t=1}^{n} \tilde{q}_t \otimes z_t]\ ,$$

and

$$R = \tilde{D}'\tilde{J}^{-1}\tilde{D}$$

$$= 3.36375 \qquad\qquad \text{(from Figure 5).}$$

Since $(\chi^2)^{-1}(.95,3) = 7.815$, the hypothesis is accepted at the 5% level.

Figure 5. Illustration of Lagrange Multiplier Test Computations with Example 1.

SAS Statements:

```
PROC MODEL OUT=MOD02;
ENDOGENOUS Y1 Y2 Y3;
EXOGENOUS  R1 R2 R3 D0 D1 D2 D3 D4 D5 D6 D7 D8 D9 D10 D11 D12 D13;
PARMS RO1 RO2 RO3 RO4 RO5 C0 C1 C2 C3 C4 C5 C6 C7 C8 C9 C10 C11 C12 C13;
T1=RO1; T2=-RO2-RO3; T3=RO2; T4=RO3; T5=RO4; T6=-RO5-RO2; T7=RO5; T8=-RO5-RO3;
PEAK= T1+T2*R1+T3*R2+T4*R3-(T2+T3+T4)*Y3;
INTER=T5+T3*R1+T6*R2+T7*R3-(T3+T6+T7)*Y3;
BASE= -1+T4*R1+T7*R2+T8*R3-(T4+T7+T8)*Y3;
Y1=LOG(PEAK/BASE);  Y2=LOG(INTER/BASE);
Y3=D0*C0+D1*C1+D2*C2+D3*C3+D4*C4+D5*C5+D6*C6+D7*C7+D8*C8+D9*C9+D10*C10+D11*C11
   +D12*C12+D13*C13;

PROC SYSNLIN DATA=EG01 MODEL=MOD02 N3SLS METHOD=GAUSS MAXIT=50 CONVERGE=1.E-8
             SDATA=IDENTITY OUTS=STILDE OUTEST=TSHARP;
INSTRUMENTS  R1 R2 R3 D0 D1 D2 D3 D4 D5 D6 D7 D8 D9 D10 D11 D12 D13 / NOINT;
FIT Y1 Y2 START = (RO1 -3 RO2 .8 RO3 .4 RO4 -1.5 RO5 .03);
PROC SYSNLIN DATA=EG01 MODEL=MOD02 N3SLS METHOD=GAUSS MAXIT=50 CONVERGE=1.E-7
             SDATA=STILDE ESTDATA=TSHARP OUTEST=RHOHAT;
INSTRUMENTS  R1 R2 R3 D0 D1 D2 D3 D4 D5 D6 D7 D8 D9 D10 D11 D12 D13 / NOINT;
FIT Y1 Y2;

DATA W01;  IF _N_=1 THEN SET RHOHAT;  SET EG01;  RETAIN RO1-RO5;
T1=RO1; T2=-RO2-RO3; T3=RO2; T4=RO3; T5=RO4; T6=-RO5-RO2; T7=RO5; T8=-RO5-RO3;
PEAK= T1+T2*R1+T3*R2+T4*R3-(T2+T3+T4)*Y3;
INTER=T5+T3*R1+T6*R2+T7*R3-(T3+T6+T7)*Y3;
BASE= -1+T4*R1+T7*R2+T8*R3-(T4+T7+T8)*Y3;
Q1=Y1-LOG(PEAK/BASE);  Q2=Y2-LOG(INTER/BASE);
DQ1T1=-1/PEAK;                       DQ2T1=0;
DQ1T2=-(R1-Y3)/PEAK;                 DQ2T2=0;
DQ1T3=-(R2-Y3)/PEAK;                 DQ2T3=-(R1-Y3)/INTER;
DQ1T4=-(R3-Y3)/PEAK+(R1-Y3)/BASE;    DQ2T4=(R1-Y3)/BASE;
DQ1T5=0;                             DQ2T5=-1/INTER;
DQ1T6=0;                             DQ2T6=-(R2-Y3)/INTER;
DQ1T7=(R2-Y3)/BASE;                  DQ2T7=-(R3-Y3)/INTER+(R2-Y3)/BASE;
DQ1T8=(R3-Y3)/BASE;                  DQ2T8=(R3-Y3)/BASE;
IF NMISS(OF D0-D13) > 0 THEN DELETE;
KEEP Q1 Q2 DQ1T1-DQ1T8 DQ2T1-DQ2T8 R1-R3 D0-D13;
```

Figure 5. (Continued).

```
PROC MATRIX;
FETCH Q1 DATA=W01(KEEP=Q1);  FETCH DQ1 DATA=W01(KEEP=DQ1T1-DQ1T8);
FETCH Q2 DATA=W01(KEEP=Q2);  FETCH DQ2 DATA=W01(KEEP=DQ2T1-DQ2T8);
FETCH Z DATA=W01(KEEP=R1-R3 D0-D13);  FETCH STILDE DATA=STILDE(KEEP=Y1 Y2);
M=J(34,1,0); DELM=J(34,8,0); V=J(34,34,0);
DO T=1 TO 220;
 QT=Q1(T,)//Q2(T,);  DELQT=DQ1(T,)//DQ2(T,);
 MT = QT @ Z(T,)';  DELMT = DELQT @ Z(T,)';
 M=M+MT;  DELM=DELM+DELMT;   V=V+STILDE @ (Z(T,)'*Z(T,));
END;
CHAT=INV(DELM'*INV(V)*DELM);  D=-CHAT*DELM'*INV(V)*M;
R=D'*INV(CHAT)*D;  PRINT R;
```

Output:

```
                              SAS                                         8

                        R                COL1

                        ROW1          3.36375
```

There is one other test that is commonly used in connection with three-stage least-squares and generalized method of moments estimation called the test of the overidentifying restrictions. The terminology is a holdover from the linear case, the test is a model specification test. The idea is that certain linear combinations of the rows of $\sqrt{n}\cdot m_n(\hat{\theta})$ are asymptotically normally distributed with a zero mean if the model is correctly specified. The estimator $\hat{\theta}$ is the minimizer of $S(\theta,\hat{V})$ so it must satisfy the restriction that $[(\partial/\partial\theta)m_n(\theta)]'\hat{V}m_n(\theta) = 0$. This is equivalent to a statement that

$$[\sqrt{n}_n m(\hat{\theta})]' - \hat{\tau}'\hat{H} = 0$$

for some full rank matrix $\hat{H}$ of order $M\cdot K - p$ by p that has rows which are orthogonal to the rows of $[(\partial/\partial\theta)m_n(\theta)]'\hat{V}$. This fact and arguments similar to either Theorem 13 of Chapter 3 or Theorem 14 of Chapter 9 lead to the conclusion that $\sqrt{n}\cdot m_n(\hat{\theta})$ is asymptotically distributed as the singular normal with a rank $M\cdot K - p$ variance-covariance matrix. This fact and arguments similar to the proof of Theorem 14 of Chapter 3 or Theorem 16 of Chapter 9 lead to the conclusion that $S(\hat{\theta},\hat{V})$ is asymptotically distributed as a chi-square random variable with $M\cdot K - p$ degrees of freedom under the hypothesis that the model is correctly specified.

One rejects the hypothesis that the model is correctly specified when $S(\hat{\theta},\hat{V})$ exceeds the upper $\alpha \times 100\%$ critical point χ^2_α of the chi-square distribution with $M\cdot K - p$ degrees of freedom; $\chi^2_\alpha = (\chi^2)^{-1}(1-\alpha,\ M\cdot K - p)$.

EXAMPLES 1 & 2. (continued). In Example 1 we have

$$S(\hat{\theta},\hat{V}) = 34.96403 \qquad \text{(from Figure 1)}$$

$$M \cdot K - p = 2 \cdot 17 - 8 = 26 \qquad \text{(from Figure 1)}$$

$$(\chi^2)^{-1}(.95,26) = 38.885$$

and the model specification is accepted. In Example 2 we have

$$S(\hat{\theta},\hat{V}) = 1.05692 \qquad \text{(from Figure 2)}$$

$$M \cdot K - p = 1 \cdot 3 - 2 = 1 \qquad \text{(from Figure 2)}$$

$$(\chi^2)^{-1}(.95,1) = 3.841$$

and the model specification is accepted.▮

PROBLEMS

1. Use Theorem 12 of Chapter 9 and the expressions given in the example of Section 5 of Chapter 9 to derive the Wald test statistic.

2. Use Theorem 15 of Chapter 9 and the expressions given in the example of Section 5 of Chapter 9 to derive the "likelihood ratio" test statistic.

3. Use Theorem 16 of Chapter 9, the expressions given in the example of Section 5 of Chapter 9, to derive the Lagrange multiplier test statistic in the form

$$R = \tilde{D}'\tilde{H}'(\tilde{H}\tilde{J}^{-1}\tilde{H}')^{-1}\tilde{H}\tilde{D}.$$

Use

$$\tilde{H}'(\tilde{H}\tilde{J}^{-1}\tilde{H}')^{-1}\tilde{H} = \tilde{J} - \tilde{J}\tilde{G}(\tilde{G}'\tilde{J}\tilde{G})^{-1}\tilde{G}'\tilde{J},$$

$$(\partial/\partial\theta)[S(\tilde{\theta},\tilde{V}) + \tilde{\mu}'h(\tilde{\theta})] = 0 \quad \text{for some Lagrange multiplier } \tilde{\mu}$$

and $\tilde{H}\tilde{G} = 0$ to put the statistic in the form

$$R = \tilde{D}'\tilde{J}^{-1}\tilde{D} .$$

5. MAXIMUM LIKELIHOOD ESTIMATION

The simplest case, and the one that we consider first, is the regression case where the errors are independently distributed, no lagged dependent variables are used as explanatory variables, and the analysis is conditional on the explanatory variables.

The setup is the same as in Section 2. Multivariate responses y_t, L-vectors, are assumed to be determined by k-dimensional independent variables x_t according to the system of simultaneous equations

$$q_\alpha(y_t,\ x_t,\ \theta_\alpha^o) = e_{\alpha t} \qquad \alpha = 1,\ 2,\ \ldots,\ L, \quad t = 1,\ 2,\ \ldots,\ n,$$

where each $q_\alpha(y,x,\theta_\alpha)$ is a real-valued function, each θ_α^o is a p_α - dimensional vector of unknown parameters, and the $e_{\alpha t}$ represent unobservable observational or experimental errors.

All the equations of the system are used in estimation so that, according to the notational conventions adopted in Section 2, $M = L$ and

$$q(y,x,\theta) = \begin{bmatrix} q_1(y,x,\theta_1) \\ q_2(y,x,\theta_2) \\ \vdots \\ q_M(y,x,\theta_M) \end{bmatrix}, \qquad e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_M \end{bmatrix}$$

$$q(y_t,x_t,\theta) = \begin{bmatrix} q_1(y_t,x_t,\theta_1) \\ q_2(y_t,x_t,\theta_2) \\ \vdots \\ q_M(y_t,x_t,\theta_M) \end{bmatrix}, \qquad e_t = \begin{bmatrix} e_{1,t} \\ e_{2,t} \\ \vdots \\ e_{M,t} \end{bmatrix}$$

where θ is a p-vector containing the non-redundant elements of the parameter vectors θ_α, $\alpha = 1, 2, \ldots, M$. The error vectors e_t are independently and identically distributed with common density $p(e|\sigma^o)$ where σ is an r-vector. The functional form $p(e|\sigma)$ of the error distribution is assumed to be known. In the event that something such as $Q(y_t,x_t,\beta) = u_t$ with $p_t(u_t) = p(u_t|x_t,\tau,\sigma)$ is envisaged, one often can find a transformation $\Psi(Q,x_t,\tau)$ which will put the model

$$q(y_t,x_t,\theta) = \Psi[Q(y_t,x_t,\beta),x_t,\tau] = \Psi(u_t,\tau) = e_t$$

into a form that has $p_t(e_t) = p(e_t|\sigma)$ where $\theta = (\beta,\tau)$.

Usually normality is assumed in applications which we indicate by writing $n(e|\sigma)$ for $p(e|\sigma)$ where σ denotes the unique elements of

$$\Sigma = \mathcal{C}(e_t,e_t')$$

viz.

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

Let $\Sigma = \Sigma(\sigma)$ denote the mapping of this vector back to the original matrix Σ. With these conventions, the functional form of $n(e|\sigma)$ is

$$n(e|\sigma) = (2\pi)^{-M/2}\det[\Sigma(\sigma)]^{-1/2}\exp[-(1/2)e'\Sigma(\sigma)^{-1}e]$$

The assumption that the the functional form of $p(e|\sigma)$ must be known is the main impediment to the use of maximum likelihood methods. Unlike the multivariate least squares case where a normality assumption does not upset the robustness of validity of the asymptotics, with an implicit model such as considered here an error in specifying $p(e|\sigma)$ can induce serious bias. The formula for computing the bias is given below and the issue is discussed is some detail the papers by Amemiya(1977,1982) and Phillips(1982).

Given any value of the error vector e from the set of admissible values $\mathcal{E} \subseteq \mathbb{R}^M$, any value of the vector of independent variables x from the set of admissible values $\mathcal{X} \subseteq \mathbb{R}^k$, and any value of the parameter vector θ from the set of admissible values $\Theta \subseteq \mathbb{R}^p$, the model

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

is assumed to determine y uniquely; if the equations have multiple roots, there is some rule for determining which solution is meant. This is the same as stating the model determines a reduced form

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

mapping $\mathcal{E} \times \mathcal{X} \times \Theta$ onto $\mathcal{Y} \subseteq \mathbb{R}^M$; for each (x,θ) in $\mathcal{X} \times \Theta$ the mapping

$$Y(\cdot,e,\theta) : \mathcal{E} \to \mathcal{Y}$$

is assumed to be one-to-one, onto. It is to be emphasized that while a reduced form must exist, it is not necessary to find it analytically or even to be able to compute it numerically in applications.

These assumptions imply the existence of a conditional density on $\mathcal{Y}$

$$p(y|x,\theta,\sigma) = |\det(\partial/\partial y')q(y,x,\theta)|\cdot p[q(y,x,\theta)|\sigma].$$

A conditional expectation is computed as either

$$\mathcal{E}(T|x) = \int_{\mathcal{Y}} T(y)\ p(y|x,\theta,\sigma)\ dy$$

or

$$\mathcal{E}(T|x) = \int_{\mathcal{E}} T[Y(e,x,\theta)]\ p(e|\sigma)\ de,$$

whichever is the more convenient.

EXAMPLE 3. Consider the model

$$q(y,x,\theta) = \begin{bmatrix} \theta_1 + \ln y_1 + \theta_2 x \\ \theta_3 + \theta_4 y_1 + y_2 + \theta_5 x \end{bmatrix}.$$

The reduced form is

$$Y(e,x,\theta) = \begin{bmatrix} \exp(e_1 - \theta_1 - \theta_2 x) \\ e_2 - \theta_3 - \theta_4 \exp(e_1 - \theta_1 - \theta_2 x) - \theta_5 x \end{bmatrix}.$$

Under normality, the conditional density defined on $\mathcal{Y} = (0,\infty)\times(-\infty,\infty)$ has the form

$$p(y|x,\theta,\sigma) = (2\pi)^{-1}(\det \Sigma)^{-1/2}(1/y_1)\exp[-(1/2)q'(y,x,\theta)\Sigma^{-1}q(y,x,\theta)].$$

where $\Sigma = \Sigma(\sigma)$.▮

The normalized, negative, log likelihood is

$$s_n(\theta,\sigma) = (1/n)\textstyle\sum_{t=1}^{n}(-1)\ln p(y|x,\theta,\sigma)$$

and the maximum likelihood estimator is the value $(\hat{\theta},\hat{\sigma})$ that minimizes $s_n(\theta,\sigma)$; that is,

$$(\hat{\theta},\hat{\sigma}) = \operatorname{argmin}_{(\theta,\sigma)} s_n(\theta,\sigma)$$

Asymptotically,

$$\sqrt{n}\begin{bmatrix}\hat\theta - \theta^o \\ \hat\sigma - \sigma^o\end{bmatrix} \xrightarrow{\mathcal{L}} N_{p+r}(0,V).$$

Put $\lambda = (\theta,\sigma)$. Either the inverse of

$$\hat{\mathcal{J}} = (1/n)\sum_{t=1}^{n}[(\partial/\partial\lambda)(-1)\ell n p(y_t|x_t,\hat\theta,\hat\sigma)][(\partial/\partial\lambda)(-1)\ell n p(y_t|x_t,\hat\theta,\hat\sigma)]'$$

or the inverse of

$$\hat{\mathcal{I}} = (\partial^2/\partial\lambda\partial\lambda')s_n(\hat\theta,\hat\sigma)$$

$$= (1/n)\sum_{t=1}^{n}(\partial^2/\partial\lambda\partial\lambda')(-1)\ell n p(y_t|x_t,\hat\theta,\hat\sigma)$$

will estimate V consistently. Suppose that $\hat V = \hat{\mathcal{J}}^{-1}$ is used to estimate V. With these normalization conventions, a 95% confidence interval on the i^{th} element of θ is computed as

$$\hat\theta_i \pm z_{.025}\sqrt{(\hat{\mathcal{J}}^{ii})}/\sqrt{n},$$

and a 95% confidence interval on the i^{th} element of σ is computed as

$$\hat\sigma_i \pm z_{.025}\sqrt{(\hat{\mathcal{J}}^{p+i,p+i})}/\sqrt{n},$$

where $\mathcal{J}^{ij}$ denotes the ij^{th} element of the inverse of a matrix $\mathcal{J}$ and $z_{.025} = N^{-1}(.025|0,1)$.

Under normality

$$s_n(\theta,\sigma) = \text{const.} - (1/n)\sum_{t=1}^{n}\ell n|\det(\partial/\partial y')q(y_t,x,\theta)| + (1/2)\ \ell n \det \Sigma(\sigma)$$

$$+ (1/2)\mathrm{tr}\left\{[\Sigma(\sigma)]^{-1}(1/n)\sum_{t=1}^{n}q(y_t,x_t,\theta)q'(y_t,x_t,\theta)\right\}$$

Define

$$J(y,x,\theta) = (\partial/\partial y')q(y,x,\theta).$$

Using the relations (Problem 4, Chapter 6)

$$(\partial/\partial\theta_i)\ell n|\det A(\theta)| = \operatorname{tr}\,[A(\theta)]^{-1}(\partial/\partial\theta_i)A(\theta),$$

$$(\partial/\partial\sigma_i)[\Sigma(\sigma)]^{-1} = -[\Sigma(\sigma)]^{-1}[\Sigma(\xi_i)][\Sigma(\sigma)]^{-1}$$

where ξ_i is the i^{th} elementary $M(M+1)/2$-vector and

$$(-1)\ell n\, p[y|x,\theta,\sigma] = \text{const.} - \ell n|\det J(y,x,\theta)| + (1/2)\,\ell n \det \Sigma(\sigma)$$
$$+ (1/2)q'(y,x,\theta)[\Sigma(\sigma)]^{-1}q(y,x,\theta)$$

we have

$$(\partial/\partial\theta_i)(-1)\ell n\, p[y|x,\theta,\sigma] = -\operatorname{tr}\{[J(y,x,\theta)]^{-1}(\partial/\partial\theta_i)J(y,x,\theta)\}$$
$$+ q'(y,x,\theta)[\Sigma(\sigma)]^{-1}(\partial/\partial\theta_i)q(y,x,\theta)$$

$$(\partial/\partial\sigma_i)(-1)\ell n\, p[y|x,\theta,\sigma] = (1/2)\operatorname{tr}\{[\Sigma(\sigma)]^{-1}\Sigma(\xi_i)\}$$
$$- (1/2)\,q'(y,x,\theta)[\Sigma(\sigma)]^{-1}[\Sigma(\xi_i)][\Sigma(\sigma)]^{-1}q(y,x,\theta).$$

Interest usually centers in the parameter θ with σ regarded as a nuisance parameter. If

$$\sigma(\theta) = \operatorname{argmin}_\sigma s_n(\theta,\sigma)$$

is easy to compute and the "concentrated likelihood"

$$s_n(\theta) = s_n[\theta,\sigma(\theta)]$$

has a tractable analytic form then alternative formulas may be used. They are as follows.

The maximum likelihood estimators are computed as

$$\hat{\theta} = \operatorname{argmin}_{\theta}\, s_n(\theta)$$

$$\hat{\sigma} = \sigma(\hat{\theta})$$

and these will, of course, be the same numerical values that would obtain from a direct minimization of $s_n(\theta,\sigma)$. Partition $\mathcal{J}$ as

$$\mathcal{J} = \begin{array}{cc} \left[\begin{array}{cc} \mathcal{J}_{\theta\theta} & \mathcal{J}_{\theta\sigma} \\ \mathcal{J}_{\sigma\theta} & \mathcal{J}_{\sigma\sigma} \end{array}\right] & \begin{array}{l} p \text{ rows} \\ r \text{ rows} \end{array} \\ \begin{array}{cc} p & r \\ \text{cols} & \text{cols} \end{array} & \end{array}$$

Partition V similarly. With these partitionings, the following relationship holds (Rao, 1973, p33.)

$$V_{\theta\theta} = (\mathcal{J}_{\theta\theta} - \mathcal{J}_{\theta\sigma}\mathcal{J}_{\sigma\sigma}^{-1}\mathcal{J}_{\sigma\theta})^{-1}.$$

We have from above that, asymptotically

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

One can show (Problem 6) that either

$$\hat{\mathcal{X}} = (1/n)\Sigma_{t=1}^{n}\Big\{(\partial/\partial\theta)\ell np[y_t|x_t,\hat{\theta},\sigma(\hat{\theta})]\Big\}\Big\{[(\partial/\partial\theta)\ell np[y_t|x_t,\hat{\theta},\sigma(\hat{\theta})]\Big\}'$$

or

$$\hat{\mathcal{I}} = (\partial^2/\partial\theta\partial\theta')s_n(\hat{\theta})$$

$$= (1/n)\ \Sigma_{t=1}^{n}(\partial^2/\partial\theta\partial\theta')(-1)\ell n p[y_t|x_t,\hat{\theta},\sigma(\hat{\theta})]$$

will estimate

$$\mathcal{J}_{\theta\theta} - \mathcal{J}_{\theta\sigma}\mathcal{J}_{\sigma\sigma}^{-1}\mathcal{J}_{\sigma\theta}$$

consistently. Thus, either $\hat{\mathcal{X}}^{-1}$ or $\hat{\mathcal{I}}^{-1}$ may be used to estimate $V_{\theta\theta}$. Note that it is necessary to compute a total derivative in the formulas above; for example,

$$(\partial/\partial\theta')(-1)\ell n p[y|x,\theta,\sigma(\theta)]$$

$$= [-1/p(y|x,\theta,\sigma)][(\partial/\partial\theta')p(y|x,\theta,\sigma)+(\partial/\partial\sigma')p(y|x,\theta,\sigma)(\partial/\partial\theta')\sigma(\theta)]\Big|_{\sigma\ =\ \sigma(\theta)}$$

Suppose that $\hat{V}_{\theta\theta} = \hat{\mathcal{X}}^{-1}$ is used to estimate $V_{\theta\theta}$. With these normalization conventions, a 95% confidence interval on the i^{th} element of θ is computed as

$$\hat{\theta}_i \pm z_{.025}\sqrt{(\hat{\mathcal{X}}^{ii})}/\sqrt{n},$$

where $\mathcal{X}^{ij}$ denotes the ij^{th} element of the inverse of the matrix $\mathcal{X}$ and $z_{.025} = N^{-1}(.025|0,1)$.

Under normality,

$$s_n(\theta,\sigma) = \text{const.} - (1/n)\Sigma_{t=1}^{n}\ell n|\det(\partial/\partial y')q(y_t,x,\theta)| + (1/2)\ \ell n \det \Sigma(\sigma)$$

$$+ (1/2)\text{tr}\left\{[\Sigma(\sigma)]^{-1}(1/n)\Sigma_{t=1}^{n}q(y_t,x_t,\theta)q'(y_t,x_t,\theta)\right\}$$

which implies that

$$\Sigma(\theta) = \Sigma[\sigma(\theta)] = (1/n)\Sigma_{t=1}^{n}q(y_t,x_t,\theta)q'(y_t,x_t,\theta)$$

whence

$$s_n(\theta) = \text{const.} - (1/n)\sum_{t=1}^{n} \ell n|\det(\partial/\partial y')q(y_t,x,\theta)|$$

$$+ (1/2)\ \ell n \det (1/n)\sum_{t=1}^{n} q(y_t,x_t,\theta)q'(y_t,x_t,\theta).$$

Using the relations (Problem 4, Chapter 6)

$$(\partial/\partial\theta_i)\ell n|\det A(\theta)| = \mathrm{tr}\ [A(\theta)]^{-1}(\partial/\partial\theta_i)A(\theta),$$

$$(\partial/\partial\theta_i)[\Sigma(\theta)]^{-1} = -[\Sigma(\theta)]^{-1}[(\partial/\partial\theta_i)\Sigma(\theta)][\Sigma(\theta)]^{-1}$$

and

$$(-1)\ell n\ p[y|x,\theta,\sigma(\theta)] = \text{const.} - \ell n|\det J(y,x,\theta)|$$

$$+ (1/2)\ \ell n \det \Sigma(\theta) + (1/2)q'(y,x,\theta)[\Sigma(\theta)]^{-1}q(y,x,\theta)$$

we have

$$(\partial/\partial\theta_i)(-1)\ell n\ p[y|x,\theta,\sigma(\theta)]$$

$$= -\ \mathrm{tr}\{[J(y,x,\theta)]^{-1}(\partial/\partial\theta_i)J(y,x,\theta)\}$$

$$+ (1/2)\mathrm{tr}\Big[[\Sigma(\theta)]^{-1}[(\partial/\partial\theta_i)\Sigma(\theta)]\{I - [\Sigma(\theta)]^{-1}q(y,x,\theta)q'(y,x,\theta)\}\Big]$$

$$+ q'(y,x,\theta)[\Sigma(\theta)]^{-1}(\partial/\partial\theta\ )q(y,x,\theta)$$

where

$$(\partial/\partial\theta_i)\Sigma(\theta) = (2/n)\sum_{t=1}^{n} q(y,x,\theta)(\partial/\partial\theta_i)q'(y,x,\theta).$$

In summary, there are four ways that one might compute an estimate $\hat{V}_{\theta\theta}$ of $V_{\theta\theta}$, the asymptotic variance-covariance matrix of $\sqrt{n}(\hat{\theta} - \theta^{o})$. Either

$\hat{\mathcal{J}}$ or $\hat{\mathcal{I}}$ may be inverted and then partitioned to obtain $\hat{V}_{\theta\theta}$ or either $\hat{\mathcal{K}}$ or $\hat{\mathcal{L}}$ may be inverted to obtain $\hat{V}_{\theta\theta}$.

As to computations, the full Newton downhill direction is obtained by expanding $s_n(\theta,\sigma)$ in a Taylor's expansion about some trial value of the parameter $\lambda_T = (\theta_T,\sigma_T)$

$$s_n(\theta,\sigma) \doteq s_n(\theta_T,\sigma_T) + [(\partial/\partial\lambda')s_n(\theta_T,\sigma_T)](\lambda - \lambda_T)$$
$$+ (1/2)(\lambda - \lambda_T)[(\partial^2/\partial\lambda\partial\lambda')s_n(\theta_T,\sigma_T)](\lambda - \lambda_T).$$

The minimum of this quadratic equation in λ is

$$\lambda_M = \lambda_T - [(\partial^2/\partial\lambda\partial\lambda')s_n(\theta_T,\sigma_T)]^{-1}[(\partial/\partial\lambda)s_n(\theta_T,\sigma_T)]$$

whence a full Newton step away from the point (θ,σ) and, hopefully, toward the point $(\hat{\theta},\hat{\sigma})$ is

$$D(\theta,\sigma) = - [(\partial^2/\partial\lambda\partial\lambda')s_n(\theta,\sigma)]^{-1}[(\partial/\partial\lambda)s_n(\theta,\sigma)].$$

A minimization algorithm incorporating partial step lengths is constructed along the same lines as the modified Gauss-Newton algorithm as discussed in Section 4 of Chapter 1. Often,

$$\mathcal{J}(\theta,\sigma) = (1/n)\sum_{t=1}^{n}[(\partial/\partial\lambda)(-1)\ell n p(y_t|x_t,\theta,\sigma)][(\partial/\partial\lambda)(-1)\ell n p(y_t|x_t,\theta,\sigma)]'$$

can be accepted as an adequate approximation to $-[(\partial^2/\partial\lambda\partial\lambda')s_n(\theta,\sigma)]$; see Problem 5.

To minimize the concentrated likelihood, the same approach leads to

$$D(\theta) = - [(\partial^2/\partial\theta\partial\theta')s_n(\theta)]^{-1}[(\partial/\partial\theta)s_n(\theta)].$$

as the correction vector and

$$\mathcal{X}(\theta) = (1/n)\sum_{t=1}^{n}\Big\{(\partial/\partial\theta)\ell np[y_t|x_t,\theta,\sigma(\theta)]\Big\}\Big\{[(\partial/\partial\theta)\ell np[y_t|x_t,\theta,\sigma(\theta)]\Big\}'$$

as an approximation to $-[(\partial^2/\partial\theta\partial\theta')s_n(\theta)]$.

As remarked earlier, the three-stage least-squares estimator only relies on moment assumptions for consistent estimation of the model parameters whereas maximum likelihood relies on a correct specification of the error density. To be precise, the maximum likelihood estimator estimates the minimum of

$$\bar{s}(\theta,\sigma,\gamma^o) = (1/n)\sum_{t=1}^{n}\int_{\mathcal{Y}}(-1)\ell np(y|x_t,\theta,\sigma)p(y|x_t,\gamma^o)dy$$

where $p(y|x,\gamma^o)$ is the conditional density function of the true data generating process by Theorem 3 of Chapter 3. When the error density $p(e|\sigma)$ is correctly specified the model parameters θ and the variance σ are estimated consistently by the information inequality (Problem 3). If not, the model parameters may be estimated consistently in some circumstances (Phillips, 1982) but in general they will not.

Consider testing the hypothesis

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

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

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

of order q by p which is assumed to have rank q at θ^o.

The Wald test statistic is

$$W = nh'(\hat{\theta})[H(\hat{\theta})\hat{V}_{\theta\theta}H'(\hat{\theta})]^{-1}h(\hat{\theta})$$

where $\hat{V}_{\theta\theta}$ denotes any of the estimators of $V_{\theta\theta}$ described above.

Let $\tilde{\theta}$ denote the minimizer of $s_n(\theta,\sigma)$ or $s_n(\theta) = s_n[\theta,\sigma(\theta)]$ subject to the restriction that $h(\theta) = 0$, whichever is the easier to compute. Let $\tilde{V}_{\theta\theta}$ denote any one of the four formulas for estimating $V_{\theta\theta}$ described above but with $\tilde{\theta}$ replacing $\hat{\theta}$ throughout. The Lagrange multiplier test statistic is

$$R = n[(\partial/\partial\theta)s_n(\tilde{\theta})]'\tilde{V}_{\theta\theta}[(\partial/\partial\theta)s_n(\tilde{\theta})] .$$

The likelihood ratio test statistic is

$$L = 2n[s_n(\tilde{\theta}) - s_n(\hat{\theta})] .$$

In each case, the null hypothesis $H\colon h(\theta^o) = 0$ is rejected in favor of the alternative hypothesis $A\colon h(\theta^o) \neq 0$ when the test statistic exceeds the upper $\alpha \times 100$ percentage point χ^2_α of a chi-square random variable with q degrees of freedom; $\chi^2_\alpha = (\chi^2)^{-1}(1-\alpha,\ q)$. Under the alternative hypothesis, each test statistic is approximately distributed as a non-central chi-square random variable with q degrees of freedom and non-centrality parameter

$$\lambda = nh'(\theta^o)[H(\theta^o)V_{\theta\theta}H'(\theta^o)]^{-1}h(\theta^o)/2$$

where $V_{\theta\theta}$ is computed by inverting and partitioning the matrix

$$\mathcal{J} = (1/n)\sum_{t=1}^{n}\int_{\mathcal{Y}}[(\partial/\partial\lambda)(-1)\ell n p(y|x_t,\theta^o,\sigma^o)][(\partial/\partial\lambda)(-1)\ell n p(y|x_t,\theta^o,\sigma^o)]'$$

$$\times\ p(y|x_t,\theta^o,\sigma^o)dy.$$

The results of Chapter 3 justify these statistical methods. The algebraic relationships needed to reduce the general results of Chapter 3 to the formulas above are given in the problems. A set of specific regularity conditions that imply the assumptions of Chapter 3, and a detailed verification that this is so, are in Gallant and Holly (1980).

In the dynamic case, the structural model has the same form as above

$$q(y_t, x_t, \theta^o) = e_t \qquad (t = 1, 2, \ldots, n)$$

with $q(y,x,\theta)$ mapping $\mathcal{Y} \times \mathcal{X} \times \Theta \subset \mathbb{R}^M \times \mathbb{R}^k \times \mathbb{R}^p$ onto $\mathcal{E} \subset \mathbb{R}^M$ and determining the one-to-one mapping $y = Y(e,x,\theta)$ of $\mathcal{E}$ onto $\mathcal{Y}$. Unlike the regression case, lagged dependent variables

$$y_{t-1}, y_{t-2}, \ldots, y_{t-\ell}$$

may be included as components of x_t and the errors e_t may be correlated. When lagged values are included, we shall assume that the data

$$y_0, y_{-1}, \ldots, y_{1-\ell}$$

are available so that $q(y_t, x_t, \theta)$ can be evaluated for $t = 1, 2, \ldots, \ell$. Let v_t denote the elements of x_t other than the lagged values.

The leading special case is that of independently and identically distributed errors where the joint density function of the errors and predetermined variables

$$p(e_n, e_{n-1}, \ldots, e_1, y_0, y_{-1}, \ldots, y_{1-\ell}, v_n, v_{n-1}, \ldots, v_1)$$

has the form

$$\prod_{t=1}^{n} p(e_t|\sigma) p(y_0, \ldots, y_{1-\ell}) p(v_n, \ldots, v_1).$$

In this case the likelihood is (Problem 1)

$$\prod_{t=1}^{n} |\det(\partial/\partial y') q(y_t, x_t, \theta) p[q(y_t, x_t, \theta)|\sigma] p(y_0, \ldots, y_{1-\ell}) p(v_n, \ldots, v_1).$$

and the conditional likelihood is

$$\prod_{t=1}^{n} |\det(\partial/\partial y')q(y_t,x_t,\theta)p[q(y_t,x_t,\theta)|\sigma].$$

One would rather avoid conditioning on the variables y_0, y_{-1}, ..., $y_{1-\ell}$ but in most applications it will not be possible to obtain the density $p(y_0, \ldots, y_{1-\ell})$. Taking logarithms, changing sign, and normalizing leads to the same sample objective function as above

$$s_n(\theta,\sigma) = (1/n)\sum_{t=1}^{n}(-1)\ell n[p(y_t|x_t,\theta)]$$

$$= (1/n)\sum_{t=1}^{n}(-1)\ell n\Big\{|\det(\partial/\partial y_t)q(y_t,x_t,\theta)|\ p[q(y_t,x_t,\theta)|\sigma]\Big\} .$$

A formal application of the results Sections 4 and 6 of Chapter 9 yields the same statistical methods as above. The algebraic relationships required in their derivation are sketched out in Problems 3 through 6.

Sometimes models can be transformed to have identically and independently distributed errors. One example was given above. As another, if the errors from $u_t = Q(y_t,v_t,\beta)$ appear to be serially correlated, a plausible model might be

$$q(y_t,x_t,\theta) = Q(y_t,v_t,\beta) - \tau Q(y_{t-1},v_{t-1},\beta) = u_t - \tau u_{t-1} = e_t.$$

with $x_t = (y_{t-1},v_t,v_{t-1})$ and $\theta = (\beta,\tau)$.

As noted, these statistical methods obtain from a formal application of the results listed in Sections 4 and 6 of Chapter 9. Of the assumptions listed in Chapter 9 that need to be satisfied by a dynamic model, the most suspect is the assumption of Near Epoch Dependence. Some results in this direction are given in Problem 2. A detailed discussion of regularity conditions for the case when $q(y,x,\theta) = y - f(x,\theta)$ and the errors are normally distributed is given in Section 4 of Chapter 9. The general flavor of the regularity conditions is that in addition to the sort of conditions that are required in the regression case, the model must damp lagged y's in the sense that for given e, v_t, and θ one should have $\|y_t\| < \|Y(e, y_{t-1}, \ldots, y_{t-\ell}, v_t, \theta)\|$.

In the general dynamic case, the joint density function

$$p(e_n, e_{n-1}, \ldots, e_1, y_0, \ldots, y_{1-\ell}, v_n, v_{n-1}, \ldots, v_1 | \sigma)$$

can be factored as (Problem 1)

$$\Pi_{t=1}^{n} p(e_t | e_{t-1}, \ldots, e_1, v_t, \ldots, v_1, y_0, \ldots, y_{1-\ell}, \sigma)$$

$$\times\, p(y_0, \ldots, y_{1-\ell}, v_n, \ldots, v_1, \sigma).$$

Put

$$p(y_t | x_t, x_{t-1}, \ldots, x_1, y_0, \ldots, y_{1-\ell}, \theta, \sigma).$$

$$= |(\partial/\partial y')q(y_t, x_t, \theta)|$$

$$\times p[q(y_t, x_t, \theta) | q(y_{t-1}, x_{t-1}, \theta), \ldots, q(y_1, x_1, \theta),\ v_t, \ldots, v_1,\ y_0, \ldots, y_{1-\ell}, \sigma).$$

Thus, the conditional likelihood is

$$p(y_n, \ldots, y_1 | v_n, \ldots, v_1, y_0, \ldots, y_{1-\ell})$$

$$= \Pi_{t=1}^{n} p(y_t | x_t, x_{t-1}, \ldots, x_1, y_0, \ldots, y_{1-\ell}, \theta, \sigma);$$

the sample objective function is

$$s_n(\theta, \sigma) = (1/n)\textstyle\sum_{t=1}^{n} (-1)\ell n[p(y_t | x_t, x_{t-1}, \ldots, x_1, y_0, \ldots, y_{1-\ell}, \theta, \sigma)];$$

and the maximum likelihood estimator is

$$(\hat{\theta},\hat{\sigma}) = \operatorname{argmin}_{(\theta,\sigma)} s_n(\theta,\sigma).$$

A formal application of the results of Section 4 of Chapter 9 yields that approximately (see Theorem 6 of Chapter 9 for an exact statement),

$$\sqrt{n}\begin{bmatrix}\hat{\theta} - \theta^o \\ \hat{\sigma} - \sigma^o\end{bmatrix} \sim N_{p+r}(0,V);$$

with

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

$\mathcal{I}$ and $\mathcal{J}$ can be estimated using

$$\hat{\mathcal{I}} = \sum_{\tau=-\ell(n)}^{\ell(n)} w[\tau/\ell(n)]\ \hat{\mathcal{I}}_{n\tau}$$

and

$$\hat{\mathcal{J}} = (\partial^2/\partial\lambda\partial\lambda')s_n(\hat{\theta},\hat{\sigma})$$

where

$$w(x) = \begin{cases} 1 - 6|x|^2 + 6|x|^3 & 0 \le x \le 1/2 \\ 2(1 - |x|)^3 & 1/2 \le x \le 1. \end{cases}$$

$$\hat{\mathcal{I}}_{n\tau} = \begin{cases} (1/n)\sum_{t=1+\tau}^{n}\left\{(\partial/\partial\lambda)\ell n[p(y_t|x_t, \ldots, x_1, y_0, \ldots, y_{1-\ell},\hat{\theta},\hat{\sigma})]\right\} \\ \quad \times\left\{(\partial/\partial\lambda)\ell n[p(y_{t-\tau}|x_{t-\tau}, \ldots, x_1, y_0, \ldots, y_{1-\ell},\hat{\theta},\hat{\sigma})]\right\}' & \tau \ge 0, \\ (\hat{\mathcal{I}}_{n,-\tau})' & \tau < 0. \end{cases}$$

For testing

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

where $\lambda = (\theta,\sigma)$ the Wald test statistic is (Theorem 12, Chapter 9)

$$W = nh'(\hat{\lambda})[\hat{H}\hat{V}\hat{H}']^{-1}h(\hat{\lambda})$$

where $\hat{H} = (\partial/\partial\lambda)h(\hat{\lambda})$. The null hypothesis $H\colon h(\lambda^o) = 0$ is rejected in favor of the alternative hypothesis $A\colon h(\lambda^o) \neq 0$ when the test statistic exceeds the upper $\alpha \times 100$ percentage point χ^2_α of a chi-square random variable with q degrees of freedom; $\chi^2_\alpha = (\chi^2)^{-1}(1-\alpha,q)$.

As a consequence of this result, a 95% confidence interval on the i^{th} element of θ is computed as

$$\hat{\theta}_i \pm z_{.025}\surd(\hat{V}_{ii})/\surd n,$$

and a 95% confidence interval on the i^{th} element of σ is computed as

$$\hat{\sigma}_i \pm z_{.025}\surd(\hat{V}_{p+i,p+i})/\surd n\ .$$

where $z_{.025} = -\surd(\chi^2)^{-1}(.95,1) = N^{-1}(.025|0,1)$.

Let $\tilde{\lambda} = (\tilde{\theta},\tilde{\sigma})$ denote the minimizer of $s_n(\lambda)$, subject to the restriction that $h(\lambda) = 0$. Let $\tilde{H}$, $\tilde{\mathcal{J}}$ and $\tilde{\mathcal{I}}$ denote the formulas for $\hat{H}$, $\hat{\mathcal{J}}$ and $\hat{\mathcal{I}}$ above but with $\tilde{\lambda}$ replacing $\hat{\lambda}$ throughout; put

$$\tilde{V} = \tilde{\mathcal{I}}^{-1}\tilde{\mathcal{J}}\,\tilde{\mathcal{I}}^{-1}.$$

The Lagrange multiplier test statistic is (Theorem 16, Chapter 9)

$$R = n[(\partial/\partial\lambda)s_n(\tilde{\lambda})]'\tilde{\mathcal{I}}^{-1}\tilde{H}'(\tilde{H}\tilde{V}\tilde{H}')^{-1}\tilde{H}\tilde{\mathcal{I}}^{-1}[(\partial/\partial\lambda)s_n(\tilde{\lambda})]\ .$$

Again, the null hypothesis $H: h(\lambda^o) = 0$ is rejected in favor of the alternative hypothesis $A: h(\lambda^o) \neq 0$ when the test statistic exceeds the upper $\alpha \times 100$ percentage point χ^2_α of a chi-square random variable with q degrees of freedom; $\chi^2_\alpha = (\chi^2)^{-1}(1-\alpha, q)$.

The likelihood ratio test cannot be used because $\mathcal{J} \neq \mathcal{I}$; see Theorem 17 of Chapter 9. Formulas for computing the power of the Wald and Lagrange multiplier tests are given in Theorems 14 and 16 of Chapter 9, respectively.

PROBLEMS

This problem set requires a reading of Sections 1 through 3 of Chapter 9 before the problems can be worked.

1. (Derivation of the likelihood in the dynamic case). Consider the model

$$q(y_t, x_t, \theta^o) = e_t \qquad (t = 1, 2, \ldots, n)$$

where

$$x_t = (y_{t-1}, y_{t-2}, \ldots, y_{t-\ell}, v_t).$$

Define the $(n+\ell)M$-vectors ζ and $e(\zeta)$ by

$$e(\zeta) = \begin{bmatrix} e_n \\ e_{n-1} \\ \vdots \\ e_1 \\ e_0 \\ \vdots \\ e_{1-\ell} \end{bmatrix} = \begin{bmatrix} q(y_n, x_n, \theta) \\ q(y_{n-1}, x_{n-1}, \theta) \\ \vdots \\ q(y_1, x_1, \theta) \\ y_0 \\ \vdots \\ y_{1-\ell} \end{bmatrix}, \quad \zeta = \begin{bmatrix} y_n \\ y_{n-1} \\ \vdots \\ y_1 \\ y_0 \\ \vdots \\ y_{1-\ell} \end{bmatrix}.$$

Show that $(\partial/\partial\zeta')e(\zeta)$ has a block upper triangular form so that

$$\det\, (\partial/\partial\zeta')e(\zeta) = \prod_{t=1}^{n} \det\, (\partial/\partial y')q(y_t, x_t, \theta).$$

Show that the joint density function

$$p(e_n, e_{n-1}, \ldots, e_1, y_0, \ldots, y_{1-\ell}, v_n, v_{n-1}, \ldots, v_1)$$

can be factored as

$$\Pi_{t=1}^{n} p(e_t | e_{t-1}, \ldots, e_1, v_t, \ldots, v_1, y_0, \ldots, y_{1-\ell})$$

$$\times\ p(y_0, \ldots, y_{1-\ell}, v_n, \ldots, v_1)$$

and hence that the conditional density

$$p(y_n, \ldots, y_1 | v_n, \ldots, v_1, y_0, \ldots, y_{1-\ell})$$

can be put in the form

$$\Pi_{t=1}^{n} p(y_t | x_t, x_{t-1}, \ldots, x_1, y_0, \ldots., y_{1-\ell}, \theta).$$

2. (Near Epoch Dependence) Consider data generated according to the nonlinear, implicit model

$$q(y_t, y_{t-1}, v_t, \theta^o) = e_t \qquad t = 1, 2, \ldots$$

$$y_t = 0 \qquad t \le 0 .$$

where y_t and e_t are univariate. If such a model is well posed then it must define y_t as a function of y_{t-1}, v_t, θ^o, and e_t. That is, there must exist a reduced form

$$y_t = Y(e_t, y_{t-1}, v_t, \theta^o) .$$

Assume that $Y(e,y,v,\theta)$ has a bounded derivative in its first argument

$$|(\partial/\partial e)Y(e,y,v,\theta)| \le \Delta$$

and is a contraction mapping in its second

$$|(\partial/\partial y)Y(e,y,v,\theta)| \le d < 1.$$

Let the errors $\{e_t\}$ be independently and identically distributed and set

$e_t = 0$ for $t \le 0$. With this structure, the underlying data generating sequence $\{V_t\}$ described in Section 2 of Chapter 9 is $V_t = (0,0)$ for $t \le 0$ and $V_t = (e_t, v_t)$ for $t = 1, 2, \ldots$. Suppose that θ^o is estimated by maximum likelihood; that is, $\hat{\theta}$ minimizes

$$s_n(\theta) = (1/n)\sum_{t=1}^{n} s(y_t, y_{t-1}, v_t, \theta)$$

$$= (1/n)\sum_{t=1}^{n} -\ln[p(y_t \mid y_{t-1}, v_t, \theta)]$$

$$= (1/n)\sum_{t=1}^{n} -\ln\left\{|\det(\partial/\partial y_t) q(y_t, y_{t-1}, v_t, \theta)| \; p[q(y_t, y_{t-1}, v_t, \theta)]\right\}$$

where $p(e)$ is the density of the errors. We have implicitly absorbed the location and scale parameters of the error density into the definition of $q(y_t, y_{t-1}, v_t, \theta)$.

Let

$$Z_t = \Delta \sum_{j=1}^{\infty} d^j |e_{t-j}|,$$

$$R_t = \sup \{|\ln p(y_t | y_{t-1}, v_t, \theta) - \ln p(y_t + h_0 | y_{t-1} + h_1, v_t, \theta)| / |h_o + h_1|\}$$

where the supremum above is taken over the set

$$A_t = \left\{(h_0, h_1, \theta) : |h_i| \le Z_{t-i}, \; \theta \in \Theta\right\}.$$

Assume that for some $p > 4$

$$\|e_1\|_p \le B < \infty$$

$$\|R_t\|_p \le B < \infty$$

Show that this situation satisfies the hypothesis of Proposition 1 of Chapter 9 by supplying the missing details in the following argument.

Define a predictor of y_s of the form

$$\hat{y}^s_{t,m} = \hat{y}_s(V_t, V_{t-1}, \ldots, V_{t-m})$$

as follows

$$\bar{y}_t = 0 \qquad\qquad t \le 0$$

$$\bar{y}_t = Y(0, \bar{y}_{t-1}, v_t, \theta^o) \qquad\qquad 0 < t$$

$$\hat{y}^s_{t,m} = \bar{y}_s \qquad\qquad s \le \max(t-m, 0)$$

$$\hat{y}^s_{t,m} = Y(e_s, \hat{y}^{s-1}_{t,m}, v_s, \theta^o) \qquad\qquad \max(t-m,0) < s \le t$$

By Taylor's theorem, there are intermediate points such that for $t \ge 0$

$$
\begin{aligned}
|y_t - \bar{y}_t| &= |Y(e_t, y_{t-1}, v_t, \theta^o) - Y(0, \bar{y}_{t-1}, v_t, \theta^o)| \\
&\le |(\partial/\partial e)Y(\bar{\bar{e}}_t, \bar{\bar{y}}_{t-1}, v_t, \theta^o)e_t + (\partial/\partial y)Y(\bar{\bar{e}}_t, \bar{\bar{y}}_{t-1}, v_t, \theta^o)(y_{t-1} - \bar{y}_{t-1})| \\
&\le d|y_{t-1} - \bar{y}_{t-1}| + \Delta|e_t| \\
&\le d^2|y_{t-2} - \bar{y}_{t-2}| + d\Delta|e_{t-1}| + \Delta|e_t| \\
&\ \ \vdots \\
&\le d^t|y_0 - \bar{y}_0| + \Delta\Sigma_{j=0}^{t-1}\, d^j|e_{t-j}| \\
&= \Delta\Sigma_{j=0}^{t-1}\, d^j|e_{t-j}|
\end{aligned}
$$

For $m > 0$ and $t - m > 0$ the same type of argument yields

$$|y_t - \hat{y}^t_{t,m}| = |Y(e_t, y_{t-1}, v_t, \theta^o) - Y(e_t, \hat{y}^{t-1}_{t,m}, v\ , \theta^o_t)|$$

$$\le |(\partial/\partial y)Y(e_t,\bar{\bar{y}}_{t-1},v_t,\theta^o)(y_{t-1}- \hat{y}^{t-1}_{t,m})|$$

$$\le d|y_{t-1} - \hat{y}^{t-1}_{t,m}|$$

$$\vdots$$

$$\le d^m|y_{t-m} - \hat{y}^{t-m}_{t,m}|$$

$$= d^m|y_{t-m} - \bar{y}_{t-m}|$$

$$= \Delta d^m \textstyle\sum_{j=0}^{t-m-1} d^j|e_{t-m-j}|$$

where the last inequality obtains by substituting the bound for $|y_t - \bar{y}_t|$ obtained previously. For $t - m < 0$ we have

$$|y_t-\hat{y}^t_{t,m}| = d^{m-t}|y_0 - y_0| = 0 \le \Delta d^m \textstyle\sum_{j=0}^{t-m-1} d^j|e_{t-m-j}|$$

In either event,

$$\|y_t - \hat{y}^t_{t,m}\|_p \le \Delta d^m \textstyle\sum_{j=0}^{t-m-1} d^j\|e_{t-m-j}\|_p \le K\Delta d^m/(1-d).$$

Letting

$$W_t = (y_t,y_{t-1},v_t)$$

$$\hat{W}^t_{t-m} = (\hat{y}^t_{t,m},\hat{y}^{t-1}_{t,m},v_t)$$

$$|a| = \textstyle\sum_{i=1}^k|a_i| \qquad \text{(for a in } \mathbb{R}^k)$$

we have

$$s_n(\theta) = (1/n)\textstyle\sum_{t=1}^n g_t(W_t,\theta)$$

with

$$g_t(W_t,\theta) = -\ell n\ p(y_t | y_{t-1}, v_t, \theta).$$

For $t \geq 1$

$$|g_t(W_t,\theta) - g_t(\hat{W}^t_{t-m},\theta)| \leq R_t |W_t - \hat{W}^t_{t-m}|$$

Letting $q = p/(p+1) < p$, $r = p/2 > 4$, and

$$B(W_t,\hat{W}^t_{t-m}) = R_t$$

we have

$$\|B(W_t,\hat{W}^t_{t-m})\|_q \leq (1 + \|R_t\|_p^p)^{1/q} \leq (1 + B^p)^{1/q} < \infty$$

$$\| B(W_t,\hat{W}^t_{t-m})\cdot|W_t - \hat{W}^t_{t-m}| \|_r \leq \|B(W_t,\hat{W}^t_{t-m})\|_{2r} \| |W_t - \hat{W}^t_{t-m}| \|_{2r}$$

$$\leq B2K\Delta d^m/(1-d) < \infty$$

$$\eta_m = \sup_t \| |W_t - \hat{W}^t_{t-m}| \|_p \leq 2K\Delta d^m/(1-d) < \infty .$$

The rate at which η_m falls off with m is exponential since $d < 1$ whence η_m is size $-q(r-1)/(r-2)$. Thus all the conditions of Proposition 1 are satisfied.

3. (Information inequality) Consider the case where the joint density function of the errors and predetermined variables

$$p(e_n, e_{n-1}, \ldots, e_1, y_0, y_{-1}, \ldots, y_{1-\ell}, v_n, v_{n-1}, \ldots, v_1)$$

has the form

$$\prod_{t=1}^n p(e_t|\sigma)p(y_0, \ldots, y_{1-\ell})p(v_n, \ldots, v_1).$$

and let

$$p(y_t|x_t,\lambda) = |\det(\partial/\partial y')q(y_t,x_t,\theta)|p[q(y_t,x_t,\theta)|\sigma].$$

Assume that $p(y|x,\lambda)$ is strictly positive and continuous on $\mathcal{Y}$. Put

$$u(y,x) = \ell n p(y|x,\lambda) - \ell n p(y|x,\lambda^o)$$

and supply the missing details in the following argument. By Jensen's inequality

$$\begin{aligned} &\exp[\int_{\mathcal{Y}} u(y,x)p(y|x,\lambda^o)dy] \\ &\quad \le \int_{\mathcal{Y}} e^{u(y,x)}p(y|x,\lambda^o)dy \\ &\quad = \int_{\mathcal{Y}} p(y|x,\lambda)dy = 1. \end{aligned}$$

The inequality is strict unless $u(y,x) = 0$ for every y. This implies that

$$\int_{\mathcal{Y}} (-1)\ell n p(y|x,\lambda)p(y|x,\lambda^o)dy > \int_{\mathcal{Y}} (-1)\ell n p(y|x,\lambda^o)p(y|x,\lambda^o)dy$$

for every λ for which $p(y|x,\lambda) \ne p(y|x,\lambda^o)$ for some y.

4. (Expectation of the score). Use Problem 3 to show that if $(\partial/\partial\lambda)\int(-1)\ell n p(y|x,\lambda)p(y|x,\lambda^o)dy$ exists then it is zero at $\lambda = \lambda^o$. Show that if Assumption 1 through 6 of Chapter 9 are satisfied then

$$\int_{\mathcal{Y}} [(\partial/\partial\lambda)(-1)\ell n p(y|x,\lambda)]p(y|\lambda^o)dy\Big|_{\lambda = \lambda^o}$$

5. (Equality of $\mathcal{J}$ and $\mathcal{I}$). Under the same setup as Problem 3, derive the identity

$$\begin{aligned} &p^{-1}(y|x,\lambda)(\partial^2/\partial\lambda_i\partial\lambda_j)p(y|x,\lambda) \\ &\quad = (\partial^2/\partial\lambda_i\partial\lambda_j)\ell n p(y|x,\lambda) + [(\partial/\partial\lambda_i)\ell n p(y|x,\lambda)][(\partial/\partial\lambda_j)\ell n p(y|x,\lambda)] \end{aligned}$$

Let ξ_i denote the i^{th} elementary vector. Justify the following steps

$$\int_{\mathcal{E}} p^{-1}[Y(e,x,\theta)|x,\lambda]\{(\partial^2/\partial\lambda_i\partial\lambda_j)p[Y(e,x,\theta)|x,\lambda]\}p(e|\sigma)de$$

$$= \ell im_{h\to 0} h^{-1}\int_{\mathcal{Y}} p^{-1}(y|x,\lambda)[(\partial/\partial\lambda_j)p(y|x,\lambda+h\xi_i)-(\partial/\partial\lambda_j)p(y|x,\lambda)]p(y|x,\lambda)dy$$

$$= \ell im_{h\to 0} h^{-1}\int_{\mathcal{Y}} p^{-1}(y|x,\lambda)[(\partial/\partial\lambda_j)p(y|x,\lambda+h\xi_i)]p(y|x,\lambda)dy$$

$$= \ell im_{h\to 0} h^{-1}\int_{\mathcal{Y}} (\partial/\partial\lambda_j)p(y|x,\lambda+h\xi_i)dy$$

$$= \ell im_{h\to 0} h^{-1}\int_{\mathcal{Y}} (\partial/\partial\lambda_j)\ell n[p(y|x,\lambda+h\xi_i)]p(y|x,\lambda+h\xi_i)dy$$

$$= \ell im_{h\to 0} h^{-1}\cdot 0 \quad = \quad 0.$$

This implies that

$$\mathcal{E}\left[p^{-1}(y_t|x_t,\lambda)(\partial^2/\partial\lambda_i\partial\lambda_j)p(y_t|x_t,\lambda)|y_0,\ \ldots,\ y_{1-\ell},v_n,\ \ldots,\ v_1)\right] = 0.$$

6. (Derivation of $\mathcal{J}$ and $\mathcal{I}$). Under the same setup as Problem 3, obtain the identity

$$0 = (\partial^2/\partial\sigma\partial\theta')s_n(\theta,\sigma)]\big|_{\sigma = \sigma(\theta)} + (\partial^2/\partial\sigma\partial\sigma')s_n(\theta,\sigma)(\partial/\partial\theta')\sigma(\theta)$$

and use it to show that

$$(\partial^2/\partial\theta\partial\theta')s_n[\theta,\sigma(\theta)]$$

$$= (\partial^2/\partial\theta\partial\theta')s_n(\theta,\sigma)\big|_{\sigma = \sigma(\theta)}$$

$$-[(\partial^2/\partial\theta\partial\sigma')s_n(\theta,\sigma)][(\partial^2/\partial\sigma\partial\sigma')s_n(\theta,\sigma)]^{-1}[(\partial^2/\partial\sigma\partial\theta')s_n(\theta,\sigma)]\big|_{\sigma = \sigma(\theta)}.$$

$$= \mathcal{I}_{\theta\theta} - \mathcal{I}_{\theta\sigma}(\mathcal{I}_{\sigma\sigma})^{-1}\mathcal{I}_{\sigma\theta}.$$

Obtain the identity

$$(\partial/\partial\theta)\ell np[y|x,\theta,\sigma(\theta)] = (\partial/\partial\theta)\ell np(y|x,\theta,\sigma)\Big|_{\sigma = \sigma(\theta)}$$

$$- \mathcal{J}_{\theta\sigma}(\mathcal{J}_{\sigma\sigma})^{-1}(\partial/\partial\sigma)\ell np(y|x,\theta,\sigma)\Big|_{\sigma = \sigma(\theta)}.$$

Approximate $\mathcal{J}_{\theta\sigma}$ and $\mathcal{J}_{\sigma\sigma}$ by $\mathcal{J}_{\theta\sigma}$ and $\mathcal{J}_{\sigma\sigma}$, and use the resulting expression to show that

$$(1/n)\Sigma_{t=1}^{n}\Big\{(\partial/\partial\theta)\ell np[y_t|x_t,\theta,\sigma(\theta)]\Big\}\Big\{(\partial/\partial\theta)\ell np[y_t|x_t,\theta,\sigma(\theta)]\Big\}'$$

$$= \mathcal{J}_{\theta\theta} - \mathcal{J}_{\theta\sigma}(\mathcal{J}_{\sigma\sigma})^{-1}\mathcal{J}_{\sigma\theta} .$$

6. REFERENCES

Amemiya, Takeshi (1977), "The Maximum Likelihood and Nonlinear Three Stage Least Squares Estimator in the General Nonlinear Simultaneous Equations Model," Econometrica 45, 955-968.

Amemiya, Takeshi (1982), "Correction to a Lemma," Econometrica 50, 1325-1328.

Anderson, T. W. (1971), The Statistical Analysis of Time Series, New York: John Wiley and Sons, Inc.

Bloomfield, Peter (1976), Fourier Analysis of Time Series: An Introduction. New York: John Wiley and Sons, Inc.

Fiacco, Anthony V. and Garth P. McCormick (1968), Nonlinear Programming: Sequential Unconstrained Minimization Techniques. New York: John Wiley and Sons, Inc.

Gallant, A. Ronald (1977), "Three-stage Least-squares Estimation for a System of Simultaneous, Nonlinear, Implicit Equations," Journal of Econometrics 5, 71-88.

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 Dale W. Jorgenson (1979), "Statistical Inference for a System of Simultaneous, Nonlinear, Implicit Equations in the Context of Instrumental Variable Estimation," Journal of Econometrics 11, 275-302.

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 26, 83-114.

Hansen, Lars P. (1985), "Using Martingale Difference Approximations to Obtain Covariance Matrix Bounds for Generalized Method of Moments Estimators," Journal of Econometrics ??, ???-???.

Hansen, Lars P. and Kenneth J. Singleton (1982), "Generalized Instrumental Variables Estimators of Nonlinear Rational Expectations Models," Econometrica 50, 1269-1286.

Hansen, Lars P. and Kenneth J. Singleton (1984), "ERRATA," Econometrica 52, 267-268.

Lucas, Robert E., Jr. (1978), "Asset Prices in an Exchange Economy," Econometrica 46, 1429-1445.

Phillips, Peter C. B. (1982), "On the Consistency of Nonlinear FIML," Econometrica 50, 1307-1324.

Rao, C. Radhakrishna (1973), Linear Statistical Inference and Its Applications, Second Edition. New York: John Wiley and Sons, Inc.

Tauchen, George (1986), "Statistical Properties of Generalized Method of Moments Estimates of Structural Parameters Using Financial Market Data," Journal of Business and Economic Statistics ??, ???-???.

7. INDEX

Across equations restrictions, 8-2-15
Autocorrelated errors, 8-4-1
Concentrated likelihood, 8-5-6, 8-5-10
Conditional likelihood, 8-5-3, 8-5-13, 8-5-15
Consumer demand, 8-1-2
Dynamic models
    defined, 8-1-1
    estimation by generalized method of moments, 8-3-1
    estimation by maximum likelihood, 8-5-1
Example 1, 8-1-2
Example 2, 8-1-5
Electricity expenditure, 8-1-2
Equation-by-equation data arrangement, 8-2-15
Estimated variance-covariance of method of moments estimators, 8-4-2
Explicit form, 8-1-1
Figure 1, 8-2-13
Figure 2, 8-3-10
Figure 3, 8-4-7
Figure 4, 8-4-11
Figure 5, 8-4-15
Functional dependency, 8-4-3
Generalized method of moments estimator
    bias of, 8-3-12
    choice of instrumental variables, 8-3-12
    confidence intervals, 8-3-12
    consistency of variance estimates, 8-3-5
    defined, 8-3-1, 8-3-5
    estimated variance-covariance of, 8-3-5
    Gauss-Newton correction vector, 8-3-6
    insuring a positive definite variance estimate, 8-3-5
    overidentifying restrictions, 8-3-12
    sample objective function, 8-3-3
Gauss-Newton correction vector, 8-2-10, 8-3-6, 8-4-2
Grouped-by-equation data arrangement, 8-2-3
Grouped-by-observation data arrangement, 8-2-3
Heteroskedasticity correction, 8-2-16, 8-2-9, 8-4-1
Implicit form, 8-1-1
Information inequality, 8-5-11, 8-5-24
Information set, 8-1-9
Instrumental variables, 8-2-6, 8-2-11, 8-2-17, 8-3-2, 8-3-6, 8-4-1
Intertemporal consumption and investment, 8-1-5
Kuhn-Tucker theorem, 8-1-10
Lagged dependent variables, 8-3-1, 8-4-1
Lagrange multiplier test
    generalized method of moments estimators, 8-4-12
    maximum likelihood estimators, dynamic structure, iid. errors, 8-5-12
    maximum likelihood estimators, general dynamic structure, 8-5-17
    maximum likelihood estimators, regression structure, 8-5-12
    three-stage least-squares estimators, 8-4-12

Likelihood ratio test
generalized method of moments estimators, 8-4-8
maximum likelihood estimators, dynamic structure, iid. errors, 8-5-12
maximum likelihood estimators, general dynamic structure, 8-5-18
maximum likelihood estimators, regression structure, 8-5-12
three-stage least-squares estimators, 8-4-8
Maximum likelihood estimator
asymptotic variance-covariance matrix of, 8-5-9
bias, 8-5-2
concentrated likelihood, 8-5-6, 8-5-10
conditional likelihood, 8-5-3, 8-5-13, 8-5-15
defined for dynamic structures with independent errors, 8-5-13
defined for general dynamic structures, 8-5-15
defined for the regression structures, 8-5-4
derivation of the likelihood in the dynamic case, 8-5-19
expectation of the score, 8-5-24
information inequality, 8-5-11, 8-5-24
Near Epoch Dependence, 8-5-14, 8-5-20
Newton downhill direction, 8-5-10
normally distributed errors, 8-5-2
sample objective function, 8-5-4, 8-5-14, 8-5-15
transformations, 8-5-2
Method of moments, 8-2-6
Mixing conditions, 8-3-2
Multivariate data arrangement, 8-2-3
Near Epoch Dependence, 8-5-14, 8-5-20
Newton downhill direction, 8-5-10
Optimal choice of instrumental variables, 8-2-17, 8-3-12
bootstrap strategy, 8-3-12
Overidentifying restrictions, 8-3-12
Parametric restriction, 8-4-2
Parzen weights, 8-3-5
Population moments, 8-2-7
Predetermined variables, 8-3-1
Rational Expectations Hypothesis, 8-1-12
Reduced form, 8-3-2, 8-5-3
Regression structure, 8-1-2, 8-3-1, 8-5-1
Robustness of validity, 8-5-2
Sample moments, 8-2-7, 8-3-3, 8-4-1
Serially correlated errors, 8-1-2, 8-3-1
Stationary, 8-3-1, 8-3-5
Three-stage least squares estimators
choice of instrumental variables, 8-2-6, 8-2-17
considered as a generalized least squares estimator, 8-2-7
correction for autocorrelation, 8-2-16
defined, 8-2-1
estimated variance-covariance matrix of, 8-2-10, 8-2-11, 8-2-16
Gauss-Newton algorithm, 8-2-10
heteroskedasticity correction, 8-2-9, 8-2-16
identification condition, 8-2-8
sample objective function, 8-2-7, 8-2-16
transformations, 8-2-2

Table 1a., 8-1-14
Table 1b., 8-1-19
Test of the overidentifying restrictions, 8-4-17
Two-stage least-squares estimator, 8-2-9, 8-2-20
Wald test statistic
    generalized method of moments estimators, 8-4-5
    maximum likelihood estimators, dynamic structure, iid. errors, 8-5-11
    maximum likelihood estimators, general dynamic structure, 8-5-17
    maximum likelihood estimators, regression structure, 8-5-11
    three-stage least-squares estimators, 8-4-5