Research Report
No. 30 1987

CONRAD

– a Maximum Likelihood Program for Estimation of Non-linear Simultaneous Equations Models

by
Erik Mellander
and Leif Jansson

THE INDUSTRIAL INSTITUTE FOR
ECONOMIC AND SOCIAL RESEARCH

The industrial Institute for Economic and Social Research

is an independent non-profit research institution, founded in 1939 by the Swedish Employers' Confederation and the Federation of Swedish Industries.

Objectives

To carry out research into economic and social conditions of importance for industrial development in Sweden.

Activities

The greater part of the Institute's work is devoted to long-term problems, especially to long-term changes in the structure of the Swedish economy particulary within manufacturing industry.

Board

Address

Industriens Utredningsinstitut
Grevgatan 34, 5 tr, S-114 53 Stockholm, Sweden
Tel. 08-783 80 00

THE INDUSTRIAL INSTITUTE FOR ECONOMIC AND
SOCIAL RESEARCH

Erik Mellander and Leif Jansson

CONRAD
- A Maximum Likelihood Program for Estimation of Non-linear Simultaneous Equations Models

Distributed by Almqvist & Wiksell International,
Stockholm, Sweden

FOREWORD

For several years the IUI has been engaged in developing non-linear estimation methods. This report, prepared by Erik Mellander in collaboration with Leif Jansson, documents some results from that work. The project was initiated by Leif Jansson and has been continued by Erik Mellander, after Leif Jansson left the institute.

Demand for the CONRAD program has grown considerably in recent years. This demand, and the tragic circumstance of Leif Jansson's death this spring, makes it appropriate to publish this intermediary report from an ongoing research program.

Stockholm, June 1987

Gunnar Eliasson

CONTENTS

1 INTRODUCTION

CONRAD (Constrained Regression with Analytical Derivatives) is a FIML[1] program for estimation of simultaneous systems of equations.

While requiring the systems to be linear in the variables, the program can handle quite general non-linearities in the parameters. The stochastic specification incorporates both contemporaneous and intertemporal correlations between the error terms, the latter in the form of a first order vector autoregression.

Owing to one of the characteristic features of CONRAD, namely the analytical determination of the gradient of the log-likelihood function, systems of considerable size can be estimated without unduly long execution times. In contrast to most other programs of comparable type, the analytical form of the gradient vector is automatically solved by the program and used to evaluate the first order derivatives. The user thus does not have to concern himself with the maximization of the likelihood function.

CONRAD can be used without prior knowledge of programming. The program offers great flexibility with respect to input and model specifications. All input may be supplied in free format. Variables can be chosen freely from a given data matrix and, regarding time series data, the estimation can be based on any coherent subperiod.

---

[1] Full Information Maximum Likelihood; for the principle of maximum likelihood, see Silvey (1975). Hendry (1976) discusses the relationship between FIML and other simultaneous equations estimates.

By means of an optional line width, printouts can be obtained both on small terminals and on full-scale printers. Not only estimation results can be output; if desired, a thorough documentation of input data and specifications will also be supplied. This facility has been implemented in such a way that it can be used as a means to check the proper arrangement of the input, before the actual estimation. In addition, a number of checks are automatically performed by the program.

Presently, CONRAD is installed on two mainframe computers, a DEC-10 at the Stockholm University Computing Center (QZ) and on a PRIME 750 at the Stockholm School of Economics.[1] Installation on other mainframe computers like e.g. IBM or Amdahl should pose only minor problem (cf. Section 3.3). The compactness of the program also makes it well suited for installation on PC's.

Finally, some reading instructions might be helpful. Users not interested in technical issues can safely skip over Section 3 and Appendix A. Those who just want to get started using CONRAD quickly, might be content to browse through Section 2.1-2.4 and Section 4 and then go on to study the example in Appendix B. The example should give a good idea about both what the program can do and how it does it. Hopefully, its length in terms of pages should not act as a deterrent - much of it is just reproductions of computer printouts.

---

[1] The installation on the PRIME was very much faciliated by the competent and imaginative help of Kjell Franzén, Stockholm School of Economics.

2 ESTIMATION AND TESTING

2.1 Formal description of the model

The simultaneous equations model is assumed to be linear in the variables. Denoting matrices and vectors by boldface type, using capitals for the matrices and small letters for the vectors, the model can be written in matrix notation in the following way

$$\mathbf{A}(\theta)\mathbf{x}_t = \mathbf{B}\mathbf{y}_t + \mathbf{C}\mathbf{z}_t = \mathbf{u}_t \qquad t=1,2,\ldots,T. \quad (1)$$

The vector $\mathbf{x}_t$ and the matrix $\mathbf{A}$ are partitioned according to

$$\mathbf{x}_t = \begin{pmatrix} \mathbf{y}_t \\ \hline \mathbf{z}_t \end{pmatrix} \quad \text{and} \quad \mathbf{A}(\theta) = (\mathbf{B}:\mathbf{C}).$$

The nx1 and mx1 vectors $\mathbf{y}_t$ and $\mathbf{z}_t$ contain observations at time t on the endogenous and predetermined variables, respectively. Lagged values of the endogenous variables may be among the explanatory variables, i.e. it is possible that $y_{t-i} \in \{\mathbf{z}_t\}$ for $i > 0$.

As indicated by the expression $\mathbf{A}(\theta)$, the elements of the nxm coefficient matrix $\mathbf{A}$ - and, accordingly, of the nxn and nxm matrices $\mathbf{B}$ and $\mathbf{C}$ - are regarded as functions of a set of k unrestricted parameters $\theta = (\theta_1, \theta_2, \ldots, \theta_k)'$. The set of all T realizations of the relationship in (1) can be compiled in the matrix equation

$$\mathbf{A}(\theta)\mathbf{X}' = \mathbf{B}\mathbf{Y}' + \mathbf{C}\mathbf{Z}' = \mathbf{U}', \quad (1a)$$

in which the row vectors $\mathbf{x}_t'$, $\mathbf{y}_t'$, $\mathbf{z}_t'$ and $\mathbf{u}_t'$ constitute the t:th rows in the corresponding matrices.

The most general stochastic specification allows for both contemporaneous and intertemporal correlation between the error terms according to:

$$\text{I:} \quad \mathbf{u}_t = \mathbf{H}\mathbf{u}_{t-1} + \mathbf{e}_t, \quad \text{where} \tag{2}$$

$$\mathbf{e}_t \sim N(\mathbf{0},\Sigma) \text{ for all } t,^1 \tag{3a}$$

such that

$$E(\mathbf{e}_s\mathbf{e}_t') = \delta_{st}\Sigma \qquad \delta_{st} = \begin{cases} 1 & \text{if } s=t \\ 0 & \text{if } s\neq t \end{cases} . \tag{3b}$$

Thus, the errors are assumed to be generated by a first order vector autoregressive process as proposed by Hendry (1971). The nxn matrix $\mathbf{H}$ is nondiagonal, making every element of $\mathbf{u}_t$ functionally dependent upon all the elements of $\mathbf{u}_{t-1}$.

Alternatively $\mathbf{H}$ can be set to the zero matrix, implying the more restrictive assumption that the stochastic process is independent of time.

In addition to I, the following assumptions are made:

II: $\mathbf{B}$ is nonsingular; det $(\mathbf{B}) \neq 0$ where "det" denotes determinant,

III: $\mathbf{X}$ has full column rank, i.e. $r(\mathbf{X}) = n+m$,

IV: Σ is unconstrained except for being posi-

---

[1] The normality assumption is not crucial. For all specifications available in CONRAD the estimate of Θ is consistent, as long as the expectation of $\mathbf{e}$ is zero and the variance is in accordance with (3b), cf. Hausman (1983).

tive definite (i.e. $\mathbf{v}'\Sigma\mathbf{v} > 0$ for all vectors $\mathbf{v} \neq \mathbf{0}$),

V: $T > n+m$,

VI: $\mathbb{H}$ is unconstrained.

Assumption II is required to ensure that the endogenous vector $\mathbf{y}_t$ be unique for every predetermined vector $\mathbf{z}_t$ and disturbance vector $\mathbf{e}_t$. Condition IV implies that (1) does not contain any identities, so it is assumed that these have been substituted out. Finally, III and V together make up a sufficient condition for the estimator to be well defined. For a more elaborate discussion about sample size requirements, see Brown (1981).

Although $\mathbb{H}$ is unconstrained, only those matrices that have all their eigenvalues within the unit circle can be considered meaningful. The eigenvalues λ_j, $j = 1,2,\ldots,n$, which are the n roots of the n:th degree equation

$$\det(\mathbb{H}-\lambda\mathbb{I}) = 0 \qquad (4)$$

where

$\mathbf{I}$ = the identity matrix of order n

λ = a+bi; a scalar with real and imaginary parts a and b, respectively (the latter possibly zero)

thus must have the property that $(a^2+b^2)^{1/2} < 1$. (In the single equation case this simplifies to the requirement that the sole element of $\mathbb{H}$ be less than one in absolute value.)

If this condition is fulfilled the process generating the errors can be taken to be stationary, which is necessary for the elements of Σ to be finite and

independent of t. When some of the eigenvalues lie outside the unit circle the error variances and covariances will instead increase without limit over time. As the program calculates the eigenvalues of the estimated **H** matrix it is possible to check which one of these cases that is prevalent.

2.2 The mapping from θ to **A**

The most general mapping from θ to an element of **A** can be considered to be the outcome of two subsequent operations.

In the first step one may apply transformation functions to the θ_i's, whose ranges are constrained to (subsets of) the non-negative or non-positive real numbers, according to

$$h_1(\theta_i) \geq \gamma_i \geq 0 \quad \text{and} \quad h_2(\theta_i) \leq \nu_i \leq 0,$$

where γ_i and ν_i are constants chosen by the user.[1] This step is treated in Section 2.2.1.

In the second step, the a_{ij}'s can be formed as a ratio of two second-degree polynomials in the transformed (or untransformed) free parameters,[2] as shown in Section 2.2.2.

---

[1] To avoid confusion it should be pointed out that h_1 and h_2 are distinct from the elements of the **H** matrix discussed in the previous section. The use of h_1 and h_2 to denote the transformation functions is due to the fact that these functions are hyperbolical.

[2] The second order polynomial mapping is considered in Jansson and Mellander (1983).

2.2.1 Inequality constraints

Often some parameters are restricted with respect to sign and/or size by the theory underlying the model. As long as the restrictions refer to individual parameters only, they can be tested by means of one-sided t-tests.

In order also to make it possible to impose inequality constraints on several parameters simultaneously,[1] the following two hyperbolic transformations have been implemented in CONRAD

$$h_1(\theta_i) = \begin{cases} (\theta_i^6 + \gamma_i^6)^{1/6} & \text{if } \gamma_i > 0 \\ (\theta_i^2 + 0.05^2)^{1/2} - 0.05 & \text{if } \gamma_i = 0 \end{cases} \tag{5a}$$

and

$$h_2(\theta_i) = -\begin{cases} (\theta_i^6 + \nu_i^6)^{1/6} & \text{if } \nu_i < 0 \\ (\theta_i^2 + 0.05^2)^{1/2} - 0.05 & \text{if } \nu_i = 0 \end{cases} \tag{5b}$$

The constant 0.05 has been used to ensure differentiability when the limit is equal to zero. As can readily be seen, the limiting values of the transformations when this constant goes to zero are $|\theta_i|$ and $-|\theta_i|$, respectively.

The user can thus impose inequality constraints on the parameters of his/her model by defining them as $h_1(\theta_i)$ or $h_2(\theta_i)$.[2]

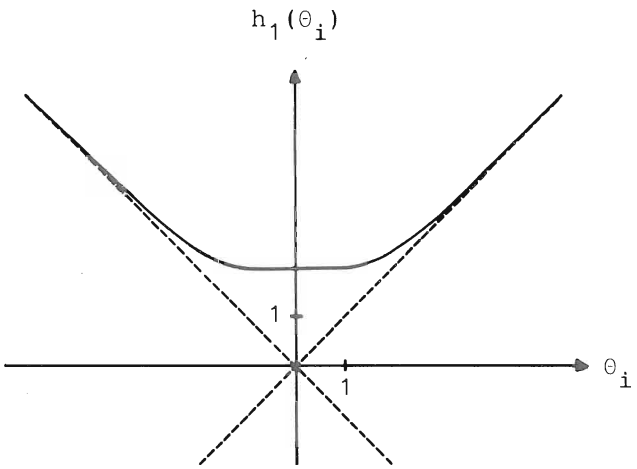
---

[1] Testing procedures for this type of constraints tend in general to be very complicated, however. See further Section 2.5.

[2] Notice the distinction between the user's model and the statistical model. The parameters of the latter are, of course, still the θ_i's.

An example of (5a) is given in Figure 1, with the lower limit set equal to 2. For illustrative purposes, 4 has been used for the exponents instead of 6. (By increasing the integer, which must be even, the hyperbola can be made to pass arbitrarily close to its asymptotes, shown by the dotted lines.)

Figure 1 The transformation function h_1 with $\gamma = 2$



The advantage of the hyperbolic transformation as compared to, e.g., a quadratic one is that the former becomes more "linear" the further away from the vertical axis one moves. So it is only in the neighborhood of the lower or upper limits (γ_i and ν_i respectively) that the restrictions are binding and any severe non-linearities are imposed. When the optimal value of θ_i is far from γ_i (or ν_i) the search for the optimal set of parameters is thus little affected. This would not be the case if a quadratic transformation had been used.

2.2.2 General form of the coefficients

To simplify the notation we write the nx(n+m) coefficient matrix **A** in the form of a column vector according to

$$\text{vec } \mathbf{A}(\theta) = \begin{bmatrix} \mathbf{a}_{.1} \\ \mathbf{a}_{.2} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{a}_{.n+m} \end{bmatrix} \qquad (6)$$

where

$\mathbf{a}_{.j}$ = the j:th column of $\mathbf{A}(\theta)$,

i.e. the columns of $\mathbf{A}(\theta)$ are stacked on top of each other. The element a_{ij} of $\mathbf{A}(\theta)$ will be in position $g = n(j-1)+i$ of vec $\mathbf{A}(\theta)$.

Further, we introduce an identity transformation of θ_i:

$$h_0(\theta_i) = \theta_i \tag{7}$$

An element of vec$\mathbf{A}(\boldsymbol{\theta})$ can then, in the most general case, be expressed in the following way:

$$a_g = \frac{n_g}{d_g}, \tag{8}$$

where

$$n_g = r_{go} + \sum_{i=1}^{k} r_{gi} \cdot h_{\ell_i}(\theta_i) + \sum_{i=1}^{k} \sum_{j \geq i}^{k} \rho_{gij} \cdot h_{\ell_i}(\theta_i) \cdot h_{\ell_j}(\theta_j) \tag{8a}$$

and

$$d_g = s_{go} + \sum_{i=1}^{k} s_{gi} \cdot h_{\ell_i}(\theta_i) + \sum_{i=1}^{k} \sum_{j \geq i}^{k} \sigma_{gij} \cdot h_{\ell_i}(\theta_i) \cdot h_{\ell_j}(\theta_j) \tag{8b}$$

$$g = 1, 2, \ldots, n(n+m)$$

$$\ell = 0, 1, 2$$

where r, ρ, s and σ denote constants.

Here, second-level indices i and j have been attached to the h_ℓ's to indicate that for a given ℓ the lower or upper limits, i.e. the γ's and the ν's, may vary with the parameters being transformed. However, since the second-level index on h is always the same as the parameter index the former will be suppressed in the following, for notational convenience.

As can be seen, the non-linear parts of (8a) and (8b) - the doublesums - contain all the k(k+1)/2 possible different products of the free parameters.

2.3 An example; the linear expenditure system

In the general form (8) the mapping from the θ_i:s to a_g looks rather complicated. In practice the mappings are usually quite simple, however, which can be seen by means of the following example.

Consider the well-known linear expenditure system [cf. Stone (1954)]. Applied to time series data, this systems of equations can be written

$$p_{it}q_{it} = \alpha_i p_{it} + \beta_i (m_t - \sum_{k=1}^{n} \alpha_k p_{kt})$$

$$= (\alpha_i - \beta_i \alpha_i) p_{it} + \beta_i m_t - \sum_{k \neq i} (\beta_i \alpha_k) p_{kt} \quad ,$$

$$i = 1 \ , \ldots , \ n$$

where p_{it} denotes the price of the i:th good in period t, q_{it} the corresponding quantity and m_t total expenditures. Total expenditures are, by definition, equal to the sum of the expenditures on the individual goods, i.e.

$$m_t = \sum_{i=1}^{n} p_{it} \, q_{it}.$$

The parameters to be estimated are the α_i's and the β_i's. As can be seen, the mappings from these parameters onto the coefficients in front of the price terms are non-linear. Further, as the β_i's have a share interpretation they should obey the following linear restriction

$$\sum_{i=1}^{n} \beta_i = 1.$$

Since we are not primarily concerned with the stochastic specification here the error terms have been suppressed. It should be mentioned, however, that they are subject to a linear restriction, which makes it necessary to drop one of the equations in the estimation. This can be seen by summing both sides of the above expenditure equation over i. Because of the relationship between individual and total expenditures and the restriction on the β_i's, both sides sum identically to m_t. Accordingly, the error terms must sum to zero. As a consequence, condition IV in Section 2.1 will not be fulfilled unless one equation is dropped. The results do not depend upon the choice of equation to be left out, however.[1]

To put the coefficients in the form of (8) consider e.g. the coefficient for p_i. If this coefficient is denoted by a_g and

$$\theta_1 \equiv \alpha_i \quad \text{and} \quad \theta_2 \equiv \beta_i$$

then

$$a_g = \theta_1 - \theta_2 \cdot \theta_1.$$

In this case the denominator (8b) is equal to unity. The program recognizes this through the setting of a parameter.[2] Accordingly, it is only necessary to specify n_g [(8a)] which is done by setting

---

[1] For a detailed discussion of this problem which is common to all types of allocation models derived from consumption theory or production theory, cf. Barten (1969). It should also be noticed that for these models intertemporal correlation between the error terms must be of a special form, see Berndt and Savin (1975), which is not available in CONRAD.

[2] ITYPE = 1; cf. Section 4.3.

$$r_{go} = 0, \quad r_{g1} = 1, \quad \rho_{g12} = -1.$$

Since the other coefficients can all be regarded as special cases of the one just considered it should be obvious how they should be specified.[1]

Had d_g [(8b)] been different from unity it would have been necessary to specify s_{go}, the s_{gi} and the σ_{gij}'s, too. Examples of such mappings can be found in Appendix B.

2.4 Identification

Whereas for linear models criteria exist for global identification,[2] the corresponding criteria for non-linear models assure only local identification, i.e. in the neighborhood of the true parameter vector.[3] However, if the non-linearities are confined to the parameters, as is the case here, the condition for local identification turns out to be quite simple.

Given the assumptions in Section 2.1, the model is locally identified if the n(n+m)xk matrix of first partial derivatives

$$\frac{\partial \mathrm{vec}\mathbf{A}(\theta)}{\partial \theta} \equiv (\psi_{gi}) = \left(\frac{\partial a_g}{\partial \theta_i}\right) \qquad \begin{matrix} g = 1,2,\ldots,n(n+m) \\ i = 1,2,\ldots,k \end{matrix} \tag{9}$$

---

[1] The linear restriction on the β_i's is simply imposed by determining the β_i in the left-out equation residually, as one minus the sum of the β_i's actually estimated.

[2] Cf. Fischer (1966).

[3] See Rothenberg (1971).

has full column rank, k. That is, the number of linearly independent columns should be equal to the number of elements in θ. If this condition is not fulfilled, the estimation algorithm will not be able to find a proper solution to the estimation problem.[1]

For most econometric models the identifiability condition is either fulfilled for all conceivable θ_i-values or for no θ_i-values at all. That is the usual, straight-forward case. However, some models may be identified for most, but not all θ_i-values. Such a situation is troublesome since in general the program cannot be prevented from assigning "bad" values to the θ_i's, i.e. values for which the matrix (9) does not have full column rank. Sometimes the inequality constraints (5a) and (5b) can be used for this purpose, however. An example of such a case is given in Appendix B.

By means of the results in Appendix A (cf. (a9)-(a13)) it can be checked analytically whether or not the model under consideration is identified for all possible θ_i-values. The necessary calculations can be rather tedious, however, especially if the number of parameters (k) is large. Further, often the only thing to do if there are some θ_i-values for which the column rank of $\partial \mathrm{vec}\mathbf{A}/\partial\theta$ is less than k, is to try to choose initial values for the parameter estimates which are not close to the ones for which the rank condition fails. Accordingly, to save time one might disregard this possible

---

[1] The characteristics of a proper solution are briefly discussed in Section 5.3, and more closely in Section 3.2. If the solution is not a proper one, it will be obvious from the content of the file RESF2.DAT that the results are useless.

problem as long as it does not arise and, if it does, first try to overcome it by arbitrarily chosing new starting values.

2.5 Hypothesis testing

The reader is assumed to be familiar with hypothesis testing relating to single parameters, i.e. test which can be carried out by means of the t-distribution.

In this section a more general test will be described, namely the likelihood ratio (LR) test. This test which, i.a., can be used in tests relating to several parameters, is shortly described in Maddala (1977, pp. 43-44).

Let H_0 be a hypothesis concerning the parametric structure of the model, which is more restrictive than an alternative hypothesis H_1. Denote the corresponding log-likelihood values by L_0 and L_1, respectively. Then $L_1 > L_0$ and minus twice the logarithm of the likelihood ratio is equal to $-2(L_0-L_1)$, which is asymptotically distributed as a chi-square (χ^2) under H_0. The number of degrees of freedom equals the difference in the number of unrestricted parameters. Formally

$$-2(L_0-L_1) \tilde{A} \chi^2(p_1-p_0) \qquad (10)$$

where

p_1 = the number of unrestricted parameters in the least restrictive model

p_0 = the number of unrestricted parameters in the more restrictive model,

and $\tilde{A}$ denotes "asymptotically distributed as". The log-likelihood values can easily be obtained according to[1]

$$L = -F + k \tag{11}$$

where

F = the function value calculated by the program

$k = -\frac{1}{2}nT[\ell n(2\Pi)+1]$

n = number of equations

T = number of observations.

However, as the constant k disappears in the subtraction of L_1 from L_0 the test can be performed directly with the F values.

Inequality constraints can in principle also be tested with the LR test. However, in the presence of such restrictions the asymptotic distribution of the test statistic is not that of a single chi-square variable but rather a weighted average of several chi-square distributions, which makes the practical application of the test very difficult in most cases. For details, see Gouriéroux, Holly and Monfort (1982) and also Judge, Griffiths, Hill and Lee (1980, Ch. 3).

Another use of the LR test is to check the compatibility of the sample information with different stochastic specifications. Denoting the more restrictive assumption of no autocorrelation with I' the following relationship can be exploited

---

[1] The corresponding formula in Appendix A is (a6). For simplicity, the superindex "*" used there, has been omitted here.

$$-2(L_I'-L_I) \underset{A}{\sim} \chi^2(n^2) \qquad (12)$$

where n is the number of equations. Note that L_I and L_I' are assumed to be based on the same number of observations.

Although the LR test is easy to use and has several attractive properties (cf. Maddala op.cit.) it has the disadvantage that it is strictly applicable only to large samples. Thus, when T is not large compared to the number of degrees of freedom, the outcomes of the test should be interpreted with care.

Finally, it should be mentioned that there are other general tests beside the LR test, e.g. the Wald and Lagrange Multiplier (LM) tests. A relatively non-technical discussion of the Wald, LR and LM tests can be found in Engle (1984). Although these tests are all asymptotically equivalent, they may yield conflicting results in finite samples. In general, the Wald test is more conservative than the LR test, while the LM test is the one most likely to reject the null hypothesis.[1]

---

[1] See Berndt and Savin (1977) and Breusch (1979).

3 PROGRAM CHARACTERISTICS

3.1 General features

CONRAD is written in FORTRAN-77. Including all subroutines the source program has a total length of about 1,600 rows, giving a rough indication of its size.

Since it has been found to considerably increase accuracy, without unduly increasing execution times, all computations are performed in double precision. To keep the price for this higher accuracy as low as possible, effort has been taken to economize on storage space. Consequently, many arrays are used for multiple purposes. Comments have, however, been inserted in the source program where this might create confusion.

Dimensioning, i.e. declarations of arrays, has been very much faciliated by use of the PARAMETER statement. As implemented in CONRAD, this statement makes it possible to redeclare all arrays in the program by changing the values of only five constants. Further, to lessen the need for re-dimensioning, in principle all matrices are stored in vector form[1], i.e. according to (6) in Section 2.2.2.

When changes none the less have to be made, the required adjustments will be indicated by the program, whereupon the execution will be automatically terminated. By means of these messages and the

---

[1] Regarding the few matrices not stored in vector form, their column indices will never have to be changed.

comments inserted after the PARAMETER statement in the main program it should be a simple matter to specify the necessary memory allocation. If, still, something should go wrong the program will print out further messages until all declarations are correct.

The use of input/output media is confined to disk files. Further, the program assumes all input to be in alphanumeric form and provides all output in the same format.

3.2 Optimization method

Since restrictions on the elements of **A** are handled by treating them as functions of the unconstrained vector θ, the maximization of the log-likelihood function is an unconstrained optimization problem. The function to be maximized is the "concentrated" log-likelihood function L* (cf. Appendix A), whose only argument is the vector θ. In practice the program instead follows the equivalent route of minimizing -L*, or, more specifically, F = -L* + k* where k* is a constant.

The minimization is carried out by means of a quasi-Newton,[1] routine, VA09AD, from the Harwell Subroutine Library,[2] which is based on an algorithm developed by Fletcher (1970). Although similar to the more well-known Davidon-Fletcher-Powell (DFP) algorithm [Fletcher and Powell (1963)], Fletcher's

---

[1] I.e. of the Newton type but not requiring evaluation of second order derivatives. See e.g. Quandt (1983, pp. 721-722).

[2] An auxiliare routine, MC11AD, from the same library is also employed.

method is considerably more efficient since it has almost eliminated the linear search subproblem inherent in the DFP.[1]

Function values and first order derivatives, which are required as input to VA09AD are provided by the subroutine MFLD. Concerning second order derivatives, only a positive definite initial estimate of the Hessian matrix is needed. We have chosen the simpliest possible alternative, i.e. the identity matrix. VA09AD then gradually approximates the inverse Hessian. The property of positive definiteness is preserved in the updating process, so as to avoid the tendency of Newton methods to move in the wrong direction when far from the optimum.

The single stopping or convergence criterion employed is the accuracy that the user requires in the estimates. Thus, the algorithm is terminated when

$$|\hat{\theta}_i^{r+1} - \hat{\theta}_i^{r}| < \text{EPS} \quad \text{for } i=1,2,\ldots,k,$$

where the superindices denote iteration number and EPS is the prespecified accuracy.[2] Although this is one obvious convergence criterion, there are several alternatives, some of which might be more efficient.[3]

---

[1] The linear search problem is discussed in Quandt (1983, pp. 734-737). For comparisons between Fletcher's algorithm and the DFP, see Fletcher (1970).

[2] Termination may also occur because the prespecified maximum number of function evaluations, MXFN, has been reached.

[3] See Belsley (1980) and Quandt (1983, pp. 737-738).

According to our experience the algorithm is very reliable. For almost any set of "reasonable" starting values for the θ_i's it converges rapidly to a local minimum.[1] In case of non-convergence the most probable cause is that the optimization problem is not well defined (for some values on the θ_i's). An illustration of such a case is given by the example in Appendix B, where the model is not identified for certain sets of parameters.

In order to avoid unnecessary long execution times it is recommended, however, that the problem is scaled such that the θ_i's do not differ in magnitude by more than a factor of 100. That can always be achieved by a slight change in the mapping from the θ_i's to the coefficients, e.g. by specifying the relevant parameters according to

$$\phi_i = c\theta_i ,$$

where c is a constant.

The first order conditions that the derivatives should be equal to zero at the terminating point are usually very closely fulfilled. When CONRAD is run on the DEC-10 computer the biggest element of the gradient vector seldom exceeds $1 \cdot 10^{-6}$ in absolute value. On the PRIME computer, which has markedly lower precision, it might happen that some derivatives are of the magnitude $1 \cdot 10^{-3}$. Other computers like e.g. IBM should be expected to fall between these two extreme values.

---

[1] The algorithm cannot discriminate between a local and a global minimum. The only way to check whether there are several minima is to try several sets of starting values and see if the terminating points differ. That is seldom the case, however.

Besides the derivatives there are two additional checks on the solution. One is the integer IEXIT, which gives the reason for the exit from VA09AD. It can take on the following values:

IEXIT=1 The normal exit, in which the accuracy condition has been fulfilled.

IEXIT=2 The specified accuracy has not been obtained. Probable cause is that EPS has been set too small for computer word length.

IEXIT=3 Maximum number of function evaluations, MXFN, has been reached. By making use of the file RESF4.DAT the search can be carried on from the point where this happened, rather than having to be started all over again (cf. Section 5.5). It should be noticed that if the model is not identified for some sets of parameter values,[1] the algorithm can in principle evaluate the function an infinite number of times in the neighborhood of these parameter values, without finding an optimum. One should thus be careful not to assign a to high value to MXFN if identification problems can be suspected.

---

[1] Cf. Section 2.4.

The other test concerns the estimate of the Hessian matrix. By construction, this matrix is positive definite and so its eigenvalues are all strictly positive. Equivalently, for the Cholesky factorization **L'dL** of the Hessian, the Cholesky values, i.e. the diagonal elements of **D**, are all positive.[1]

One (or several) of the Cholesky values may be very small, however, indicating that the solution is not wholly satisfactory. That can happen, for instance, if one of the parameters is just barely identified, due to the strong collinearities in the data. In such cases the difference between the largest and the smallest of the Cholesky values will be big and so the ratio.

$$\delta = d_{min}/d_{max}$$

will be small, where d_{min} is the smallest of the Cholesky values and d_{max} the largest.

One measure that can be used as a lower bound for δ is the Euclidean norm of the gradient vector. Denoting the gradient vector by **g**, the Euclidean norm is defined according to

$$\|\mathbf{g}\| = \left(\sum_{i=1}^{k} g_i^2 \right)^{1/2}$$

Thus, for a well-behaved solution the following inequality should hold

$$1 > \frac{\|\mathbf{g}\|}{\delta} = \text{COND},$$

---

[1] The Cholesky factorization is briefly discussed in Theil (1983). An extensive treatment and a proof of the stated equivalence can be found in Lau (1978).

COND being the number computed by the program.

One problem with this measure is that it is dependent upon the degree to which the first order conditions are fulfilled. The further away from zero the first order derivatives are the larger will be the norm of **g** and, accordingly, COND. When CONRAD is run on a relatively imprecise machine like the PRIME computer COND will accordingly be higher than otherwise, sometimes failing to fulfill the criterion. (Indirectly, δ will also depend upon the derivatives, through the construction of the optimization algorithm. However, with respect to this constant it is impossible to know the direction of the effect.)

Before concluding this section an interesting result reported by Belsley (1980) will be remarked upon.

Considering the DFP algorithm, Belsley finds that execution times can be considerably reduced (up to 50 percent) if the Hessian is initialized by the estimate proposed by Berndt, Hall, Hall and Hausman (1974), instead of the identity matrix. The fact that the BHHH Hessian is quite easy to program makes the suggested approach particularly attractive.

According to a (very) small number of test, Belsley's conclusions do not seem to hold for Fletcher's algorithm, however. In fact, initialization by the BHHH Hessian actually increased execution times slightly in a few cases. The procedure has thus not been implemented in the present program.

3.3 Computation of Var($\hat{\theta}$)

It can be shown that the variance of $\hat{\theta}$ fulfills the following inequality[1]

$$\mathrm{Var}(\hat{\theta}) \geq -\{E[\frac{\partial(\partial L^*(\theta_o)/\partial\theta)'}{\partial\theta}]\}^{-1} \equiv \mathbf{Q}^{-1}$$

where θ_o is the true parameter vector and $\mathbf{Q}$ the so called information matrix. Since matrices are involved, the interpretation of the inequality sign is that the difference between Var($\hat{\theta}$) and $\mathbf{Q}^{-1}$ is a positive semidefinite matrix. Equality holds asymptotically, as the sample size approaches infinity.

One way to estimate Var($\hat{\theta}$) is thus to determine the analytic expression for $\mathbf{Q}^{-1}$ and evaluate it at $\hat{\theta}$. For the model in Section 2, $\mathbf{Q}$ can be found by a simple generalization of Hendry's (1974) result, so as to include nonlinear restrictions on the parameters.[2]

However, since $\mathbf{Q}$ is a rather complicated matrix, which is difficult to program, we have instead based our estimate of Var($\hat{\theta}$) on the Hessian matrix produced by the minimization routine (cf. above). A justification for this is given by the following equality,[3]

$$\mathbf{Q}^{-1} = -\mathrm{plim}[\frac{\partial(L^*(\hat{\theta})/\partial\theta)'}{\partial\theta}]^{-1} \equiv -\mathrm{plim}[\mathbf{G}(\hat{\theta})]^{-1}$$

---

[1] See Silvey (1975).

[2] See Mellander (1984).

[3] Cf. Pollock (1979, p. 345).

where plim denotes probability limit and $\mathbf{G}(\hat{\theta})$ the Hessian, evaluated at $\hat{\theta}$. The inverse of $\mathbf{G}(\hat{\theta})$ is thus a consistent estimate of the lower bound for $\mathrm{Var}(\hat{\theta})$.

As the minimization algorithm makes a second order approximation of the function to be minimized, the Hessian matrix obtained from it will be equal to $\mathbf{G}(\hat{\theta})$ only if L* is quadratic. However, at least in large samples, L* is known to be very nearly so. Further, for the models considered by Belsley (1980) the Hessian matrices produced by two different quasi-Newton routines were very close to their analytic counterparts in spite of the sample sizes being very moderate, 20-40 observations.

Although the two estimates of $\mathrm{Var}(\hat{\theta})$ described above are asymptotically equal they may of course differ in small samples. In general, variance estimates based on the Hessian matrix tend to be greater than those based on the information matrix.[1] Accordingly, one would expect inferences based on the former estimates to be more cautious if the inferences are based on t-tests.[2] Inferences based on likelihood ratio tests will not be affected, however.

3.4 Compatibility with other computers

Since CONRAD has been written in accordance with standard FORTRAN-77 conventions it should be easy

---

[1] Cf., e.g. Calzolari and Panattoni (1983).

[2] More cautions in the sense of reducing the probability of so called type I error, i.e. of wrongly rejecting the null hypothesis when it is true.

to implement on other computers than DEC-10 and PRIME 750. Probably, the only obstacle concerns the input and output of data and the handling of the corresponding files. The subroutine OC which opens and closes input and output files may not be directly transferable to other computers. Further, terminal output is not generated in the same way on different computers. On the DEC machine terminal output is obtained by means of the TYPE statement, whereas on an IBM computer it would be generated by routing the output to the device with logical unit number 5.

4 INPUT INSTRUCTIONS

Three input files are used. Each of these will be described separately, under headings equal to the names which should be assigned to them.

Common to all the three files is the use of free format input. Hence, considering a particular row the only requirement is that the strings (values) appear in the prescribed order, separated by at least one blank position. There is only one exception to this rule and it concerns the naming of variables, cf. Section 4.3.

Regarding the observations on the endogenous and predetermined variables, a distinction is made between the dataset in the input file and the data actually used in the estimation. The latter may be a subset of the former, differing both in number of variables and/or observations.

The choice of variables is completely free. Arbitrary subsystems of the total system comprised by the data material may thus be estimated without renewing the data input. The choice of observations is somewhat more limited, but, e.g., in the context of time series data it allows the estimation to be based on any coherent subperiod.

Sections 4.1, 4.2 and 4.3 all contain schematic representations of the corresponding files. These are such that unless otherwise specified one row in the manual corresponds to one row in the file. The notations have been chosen so as to be easily recognizable with those used in Section 2.2.

When the input files have been properly arranged the execution can be started by the command RUN CONRAD. If the program discovers any errors in the input data or if the memory allocation is insufficient for the problem at hand instructions about how to make corrections will automatically be printed out, whereupon the execution will be deleted.

4.1 DATAM.DAT

This file contains observations on the endogenous and predetermined variables.

The data must be arranged in matrix form. The matrix should be structured so that the endogenous and predetermined variables form block matrices, according to

$$\begin{matrix} y_{11} & y_{12} & \cdots & y_{1\eta} & z_{11} & z_{12} & \cdots & z_{1\varsigma} \\ y_{21} & y_{22} & \cdots & y_{2\eta} & z_{21} & z_{22} & \cdots & z_{2\varsigma} \\ \cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot \\ y_{\tau 1} & y_{\tau 2} & \cdots & y_{\tau\eta} & z_{\tau 1} & z_{\tau 2} & \cdots & z_{\tau\varsigma} \end{matrix}$$

where

η = the number of endogenous variables,
ς = the number of predetermined variables,
τ = the number of observations.

As mentioned in the previous section, the data actually used in the estimation may be a subset of the data matrix in this file. The indices above thus constitute upper limits for the corresponding

indices of the **X** matrix (cf. Section 2.1), according to $n \leq \eta$, $m \leq \varsigma$ and $T \leq \tau$.

In principle, one row in the data matrix should correspond to one row in the input file. However, with many variables that may be impossible, due to lack of space in the file's rows. In that case the matrix can be partioned into submatrices of appropriate sizes. The submatrices should then be "stacked" in the file, one underneath the other. For instance, if the user wants to split up the data in two parts, the first part containing all the endogenous variables and the first predetermined variable, that could be done in the following way:

$$
\begin{array}{cccccc}
y_{11} & y_{12} & \cdots & y_{1\eta} & z_{11} \\
y_{21} & y_{22} & \cdots & y_{2\eta} & z_{21} \\
\cdot & \cdot & & \cdot & \cdot \\
\cdot & \cdot & & \cdot & \cdot \\
\cdot & \cdot & & \cdot & \cdot \\
y_{\tau 1} & y_{\tau 2} & \cdots & y_{\tau\eta} & z_{\tau 1}
\end{array}
$$

$$
\begin{array}{cccc}
z_{12} & \cdots & z_{1\varsigma} \\
z_{22} & \cdots & z_{2\varsigma} \\
\cdot & & \cdot \\
\cdot & & \cdot \\
\cdot & & \cdot \\
z_{\tau 2} & \cdots & z_{\tau\varsigma}
\end{array}
$$

The number of blocks and the number of columns in each block is freely determined by the user. Different blocks need not have equally many columns. They must, however, have the same number of rows (i.e. τ).

The blocks may be separated by blank rows. This can be convenient, e.g. by making it easier to check the file. The number of blank rows - which may be zero - is set by the user. However, all blocks must be separated by the same number of blank rows. No blank rows should preceed the first block.

4.2 PARIC.DAT

Initial values for the unrestricted parameters and information about inequality constraints should be provided in this file.

File structure

k			
$\bar{\theta}^1$	$i(\bar{\theta}^1)$	$\ell(\bar{\theta}^1)$	$\gamma/u(\bar{\theta}^1)$
$\bar{\theta}^2$	$i(\bar{\theta}^2)$	$\ell(\bar{\theta}^2)$	$\gamma/u(\bar{\theta}^2)$
.	.	.	.
.	.	.	.
.	.	.	.
$\bar{\theta}^k$	$i(\bar{\theta}^k)$	$\ell(\bar{\theta}^k)$	$\gamma/u(\bar{\theta}^k)$

where

k — the number of unrestricted parameters to be estimated,

$\bar{\theta}^\alpha$ — $\alpha=1,\ldots,k$. The initial value of the parameter having the α:th ordinal number. (Ordinal numbers will always be indicated by superindices, so as to avoid confusion with the subindices used in Section 2.)

$i(\bar{\theta}^\alpha)$ — $1 \leq i \leq k$. The (sub)index of the parameter whose initial value is $\bar{\theta}^\alpha$.

$\ell(\bar{\theta}^{\alpha})$ =0 if no hyperbolic transformation should be applied to the parameter. The 0 should be followed by (at least) one blank position and a "/" (slash).
=1 if the parameter should be transformed according to (5a).
=2 if the parameter should be transformed according to (5b).

$\gamma/\upsilon(\bar{\theta}^{\alpha})$ Specified only if $\ell=\ell(\bar{\theta}^{\alpha})\geq 1$. If $\ell=1$ the parameter's lower limit, i.e. γ, should be given and if $\ell=2$ the upper limit, υ.

In general it does not matter very much how the initial values are constructed. The program will almost always find a solution if they are not wholly unreasonable. However, the execution time can be considerably reduced if the starting values are based on some kind of consistent estimates.

For a system of the "seemingly unrelated" type (Zellner (1962)) OLS provides consistent estimates. The linear expenditure system considered in Section 2.3 is of this kind. However, like all allocation models that model has a characteristic feature which somewhat complicates the application of OLS to generate initial values. The common characteristic is that some or all of the parameters appear in every equation of the system. Hence, OLS estimation of the individual equations will provide

several estimates of one and the same parameter.[1] Usually, it will work well to use the mean values of the various OLS estimates as starting values.

For a (truly) simultaneous equations model the most simple consistent estimator is given by 2SLS (two-stage least squares), applied without taking the non-linear restrictions on the parameters explicitly into account. An example of this technique is given in Appendix B.

4.3 GICOF.DAT

In this file some general information should first be given, regarding, i.a., the stochastic specification and the format of output. Then the file DATAM.DAT should be described. Finally, the desired data set and the nonzero coefficients have to be specified.

A subset of the data in DATAM.DAT is determined in the following way. The relevant observations are indicated by the row numbers of the data matrix corresponding to the "first" and "last" observations. To select the appropriate variables, maximum numbers are first set for both the endogenous and the predetermined variables. In the respective block matrices (cf. 4.1), these numbers are equal to the column numbers corresponding to the rightmost of the considered variables. Whether the

---

[1] Of course, if the equations are non-linear in the parameters OLS will not provide estimates of all parameters directly. Regarding, e.g., the linear expenditure system OLS will yield estimates of $(\alpha_i - \beta_i\alpha_i)$, β_i and $\beta_i\alpha_j$, $j \neq i$, respectively. From these expressions all the unknown parameters can easily be solved for, however.

maximum numbers of variables will be equal to or greater than the number of variables actually taken into account in the estimation is decided by the specification of the coefficients in the **A** matrix.

All coefficients not explicitly specified will be set equal to zero. This also applies to the diagonal elements of the **B** matrix (which usually are automatically set to -1 in most program). Among the variables chosen in the first step (cf. above), certain ones can thus be ignored in the estimation by not specifying the coefficients associated with them.

No difference is made between θ_i, $h_1(\theta_i)$ and $h_2(\theta_i)$ when the coefficients are specified. All three of these expressions are identified by the index i. The task here is thus only to assign values to the constants in the polynomials (8a) and (8b).

As all nonzero coefficients should be specified according to the same principle only one coefficient is shown in the schematic file structure below. For easy reference to (8), constants in the polynominal have been written with small letters. To simplify the notation the common index g has been suppressed, however.

As in the previous section, superindicies denote ordinal numbers.

File structure

LP EPS MXFN WIDTH HEAD
NY NZ NO
NBLC ICL(1) ... ICL(NBLC) NBR
IFO ILO NMX MMX
(Variable names)
(Title)
NA IOR^1 IOR^2 . . .
IA JA ITYPE

r_o	ℓn	$n\ell n$	s_o	ℓd	$n\ell d$	
r^1		$i[r^1]$				
.		.				
.		.				
$r^{\ell n}$		$i[r^{\ell n}]$				
ρ^1		$i[\rho^1]$		$j[\rho^1]$		
.		.		.		
.		.		.		
$\rho^{n\ell n}$		$i[\rho^{n\ell n}]$		$j[\rho^{n\ell n}]$		a_g
s^1		$i[s^1]$				
.		.				
.		.				
$s^{\ell d}$		$i[s^{\ell d}]$				
σ^1		$i[\sigma^1]$		$j[\sigma^1]$		
.		.		.		
.		.		.		
$\sigma^{n\ell d}$		$i[\sigma^{n\ell d}]$		$j[\sigma^{n\ell d}]$		

where

LP Determines the stochastic specification:
=1 for autocorrelated errors
=0 otherwise.

EPS A real positive number, specifying the accuracy of the estimates (cf. 3.2).

MXFN Should be set to 0 in the first run, resulting in printout of all input data in RESF1.DAT. Otherwise set equal to the maximum number of evaluations of the log-likelihood function.

WIDTH The maximum number of character positions per row in the main output file.

HEAD If the user wants to provide the output files with a title, then HEAD=1, otherwise HEAD=0.

NY The number of endogenous variables in the data matrix in DATAM.DAT.

NZ The number of predetermined variables in the data matrix in DATAM.DAT.

NO The number of observations on the variables in DATAM.DAT.

NBLC The number of blocks into which the data matrix in DATAM.DAT is partitioned. If NBLC=1 the "1" should be followed by (at least) one blank position and a "/" (slash).

ICL(i) Specified only if NBLC>1. The number of columns in the i:th block, i=1,2,..., NBLC.

NBR Specified only if NBLC>1. The number of blank rows between the blocks. Assumed to be zero if omitted.

IFO, ILO The row numbers of the data matrix in DATAM.DAT, which correspond to the first and last observations, respectively, that are to be considered in the estimation.

NMX MMX The maximum numbers of endogenous and pre-determined variables, respectively, that are to be considered in the estimation (cf. the beginning of this section).

Variable names The NMX+MMX variable names, given in the same order that the variables appear in the **X** matrix. Each name can contain a maximum of 8 characters, including blanks. The maximum number of names per line is 9, the first 8 positions being reserved for the first name, positions 9 to 16 for the second name and so on. If NMX+MMX > 9 the 10:th to 18:th names are given on the second line, the 19:th to 27:th on the third line and so forth.

Title Given only if HEAD=1. The maximum length of the title is 72 characters, including blanks.

NA The number of <u>all</u> nonzero coefficients in the **A** matrix.

IOR^i The numbers of the equations containing intercepts. These are required for the calculation of goodness-of-fit measures. A particular equation is identified by the number of the column in the data matrix that contains its endogenous variable. The last number should be followed by (at least) one blank position and a "/" (slash).

IA,JA The coefficients' row and column indices, respectively, in the **A** matrix.

ITYPE =0 if the coefficient is restricted, a priori, to be equal to a constant,
=1 if the coefficient is equal to (a special case of) n_g in (8a).
=2 otherwise.

r_o The constant in the polynomial n_g, which constitutes the numerator of a_g. If - and only if - ITYPE=0, r_o is set equal to the a priori constant, followed by (at least) one blank position and a "/" (slash), thereby completing the specification of the coefficient.

ℓn, $n\ell n$ Specified only if ITYPE>0. The number of terms in the linear and nonlinear parts of n_g, respectively, Thus, ℓn is equal to the number of r's and $n\ell n$ to number of ρ's. Either of these might be equal to zero but not both at the same time. If - and only if - ITYPE=1 then "$n\ell n$" should be followed by (at least) one blank and a "/" (slash).

s_o, ℓd Specified only if ITYPE=2. The s_o is the constant in d_g. ℓd and $n\ell d$ are defined in analogy with ℓn and $n\ell n$, i.e. ℓd is the number of s's and $n\ell n$ the number of σ's.

r^κ, $i[r^\kappa]$ Specified only if $\ell n > 0$. $\kappa = 1, \ldots, \ell n$. r^κ is the κ:th constant and $i[r^\kappa]$ the index of the parameter associated with this constant. (Notice the difference between k = the number of parameters to be estimated[1] and the greek letter κ used here). As indicated by the use of superindices, no particular order is required.

ρ^μ, $i[\rho^\mu]$, $j[\rho^\mu]$ Specified only if $n\ell n > 0$. $\mu = 1, \ldots, n\ell n$. The notation is analogous with that used for the r's. The i- and j-indices may be equal.

If ITYPE=2, the nonzero s- and σ-constants with related indices are then specified, in the same way as the r's and σ's.

---

[1] Cf. Section 4.2.

5 COMMENTS ON OUTPUT

The results are presented in four output files. In addition, some information - e.g. concerning the development of the likelihood function in the search for the optimum - is provided as terminal output, during the execution.

The content of this section is organized in the same way as that of the preceding one, i.e. it is divided into subsections corresponding to the different output files. Further, there is also a short subsection about the terminal output.

It is recommended that this section and the example in Appendix B be read in parallel.

5.1 RESF1.DAT

This file will be created only if the parameter MXFN is set to zero. In that case the input to CONRAD, organized in easily readable form, will be printed here, whereupon the execution will be terminated. This should enable the user to check that the input has been properly arranged, before turning to the actual estimation. Another advantage is that this file provides a thorough documentation of the data and specifications used in the regression.

Concerning the data, only the NMX and MMX first columns of the **Y** and **Z** matrices are printed, respectively, beginning with the IFO:th observation and ending with the ILO:th. (The format is F11.6.)

The number of observations (i.e. ILO-IFO+1) is then explicitly given, together with the values of IFO and ILO.

If the vector autoregressive stochastic specification has been chosen, that will be indicated by a statement saying that the number of observations on the model is one less than the number of observations on the variables.

The initial values for the parameters are then given, together with information about the inequality constraints. Actually, the initial values subject to the chosen hyperbolic transformations are given, rather than the initial values themselves. The reason for this is just that it makes it easier to check the equality constraints (cf. below).

Since the program sorts the rows of PARIC.DAT according to parameter index, the initial values and information about the transformations will appear in increasing order, irrespective of the order in which they have been input. The transformations are identified by the numbers 0 (no transformation), 1 or 2 as set out in Section 2.2.

The required accuracy in the estimates, i.e. the value of the parameter EPS (cf. Section 3.2), is also shown.

Finally, the values of the nonzero elements in the coefficient matrix **A** are given, together with the corresponding equality constraints. The values of the a_{ij}'s should be the ones implied by the specified constraints and the initial values of the (hyperbolically transformed) parameters. No change is made with respect to order, i.e. the coeffici-

ents appear in the same order as they have been input in GICOF.DAT.

For each of the coefficients the type of constraint is indicated by the value of ITYPE (cf. Section 4.3). The constants in the mapping from the (transformed) parameters are also given, together with the indices of the corresponding parameters. The constants are denoted in the same way as in (8a) and (8b) in Section 2.2.2, except that the ρ's and σ's are denoted "greek r" and "greek s", respectively.

5.2 Terminal output

One type of terminal output concerns checks on input and problem specification. The program performs a number of such checks, before starting the actual estimation. If an error is detected it is communicated to the user via the terminal, whereafter the execution is stopped.

During the optimization the function value and its two basic components will be printed according to

$$F \quad \ell n[\det(\mathbf{B})] \quad \ell n[\det(\Sigma)]$$

each time the function is evaluated.[1] If the model is specified either in the form of a single equation or as a system of "seemingly unrelated" equations (like, e.g., the linear expenditure system considered in Section 2.3) $\ell n[\det(\mathbf{B})]$ will be identically equal to zero. The relationship between the function F and the two determinants is given by

---

[1] ("ℓn" and "det" denote natural logarithm and determinant, respectively).

$$F = \frac{1}{T}\left[\frac{1}{2}\ell n(\det(\Sigma)) - \ell n(\det(\mathbf{B}))\right]$$

whereas the relationship between F and the log-likelihood function is given in Section 2.5.

After the final iteration, "END OF VA09" will be printed. The estimated parameter vector and the corresponding gradient vector are then given, their elements ordered according to parameter index. A final printing of F and the logarithms of the two determinants conclude the terminal output.

5.3 RESF2.DAT

This is the main output file. The first line contains some technical information, conveyed through the variables IEXIT and COND. If the run has been successful IEXIT should equal unity and COND be less than 1 (cf. Section 3.2). The final F-value and the logarithms of the corresponding determinants are then given; cf. the previous section.

To enable computation of the generalized R^2 measure proposed by Berndt[1] the logarithm of the determinant of a matrix denoted (**Y-YBAR**)'(**Y-BAR**) is also given. The generalized R^2, which measures the goodness of fit for the whole system of equations, can be obtained according to

---

[1] Cf. Berndt and Khaled (1979, p. 1228). Other generalized measures have been proposed by e.g. Carter and Nagar (1977). The advantage of Berndt's measure is that it is the only one which is invariant to whether the structural or the reduced form of the system is considered.

$$\tilde{R}^2 = 1 - \exp\{\ell n(\det(\Sigma)) - 2\ell n(\det(\mathbf{B}))$$

$$- \ell n[\det((\mathbf{Y-BAR})'(\mathbf{Y-YBAR}))]\}.$$

Ordinarily, the i:th column of **YBAR** will contain the mean value of the i:th endogenous variable, the matrix **Y-BAR** thus containing the deviations of the endogenous variables from their respective means. However, if the i:th equation does not contain an intercept, the i:th column of **YBAR** will contain zeros only.

In case some of the NMX first equations are not considered in the estimation they are of course not considered in the calculation of det[(**Y-YBAR**)' (**Y-YBAR**)] either.

To indicate the relevant alternative of these three possibilities with respect to the different equations, a row of the matrix **YBAR** has been printed below the logarithm of det[(**Y-YBAR**)'(**Y-YBAR**)]. Accordingly, for the equations considered in the estimation, the i:th element in the row will either be equal to the mean value of the endogenous variable in the i:th equation or have a zero value, depending on whether the equation contains an intercept or not. Equations which have not been considered in the computation of the determinant are indicated by the value -999.

The estimated parameter vector is then given, followed by the corresponding vectors of first order derivatives (the gradient) and standard errors, respectively. The elements of the gradient vector should be close to zero, cf. Section 3.2, p. 27.

To make it possible to calculate the variances of coefficients involving several parameters, the complete covariance matrix of the parameter vector is also printed.[1] (The standard errors have, of course, been obtained as the square root of the diagonal elements of this matrix.) The heading, HESIAN(-1), where (-1) indicates inverse, refers to the way this matrix has been estimated; see further Section 3.3.

Most of the following output should be easily understood, since the notation corresponds directly to the one used in Section 2.

The goodness of fit measure used for the individual equations needs some explanation, however. It is equal to the squared cosine of the angle between either the vector of observations on the endogenous variable (**y**) and the corresponding predicted vector (**yp**), or between the same vectors, measured as deviations from their respective means.[2] The latter variant, which is used if the equation contains an intercept, is equal to the squared correlation coefficient between **y** and **yp**. [Cf. Wonnacott and Wonnacott (1970, p. 303).]

In contrast to the conventional R^2, which is not bounded from below when applied to one of the equations in a simultaneous system, this measure always lies in the closed interval [0,1]. It is

---

[1] General formulas for such computations can be found in Klein (1953, p. 258) or Fomby, Hill and Johnson (1984, p. 58).

[2] Cf. Haessel (1978).

also applicable in the case of non-linearities, which R^2 is not.[1]

The d statistic of Durbin and Watson (1950, 1951) is supplied in connection with the goodness of fit measure (under the heading "D-W"). It should be pointed out, however, that the distributional properties of this statistic are not well known in the context of simultaneously estimated systems of equations.[2] This holds in particular if non-linearities are present. It is thus intended as an indicative measure rather than as a test statistic. For testing purposes the LR test described at the end of Section 2.5 is instead recommended.[3]

5.4 RESF3.DAT

The content of this file is supplied by the optimization algorithm. Presumably, it will be of interest only if the execution is unsatisfactory in some technical sense - as indicated, e.g., by the values of IEXIT and COND, given at the beginning of RESF2.DAT.

---

[1] There is a relationship between the squared cosine measure and the ordinary R^2, however, which makes them take on identical values when applied to a linear equation with intercept, estimated by OLS.

[2] Cf., however, Durbin (1957).

[3] If, still, used as a test statistic Farebrother's (1980) tables should be used if the equation does not contain an intercept. Further, if the equation contains lagged endogenous variables the "h" statistic should be substituted for the d statistic. The former can, however, be computed with the help of the latter, cf. Durbin (1970, p. 419).

For every fifth iteration, and also on exit, the following is printed

Iteration number, Number of calls on MFLD
The value of F

$\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k$ (5 to a line)
$g_1, g_2, \ldots, g_k$ (- " -)

where g_i is the first order derivative of F with respect to $\hat{\theta}_i$.

On exit, the value of IEXIT is also printed, after the number of calls on MFLD.

It is the number of calls on MFLD that is delimited by the setting of the parameter MXFN (cf. Section 4.3). The number of iterations will be at most equal to the number of calls on MFLD but usually somewhat smaller.[1]

5.5 RESF4.DAT

This file is organized in the same way as the input file PARIC.DAT, the only difference being that the final estimates of the parameters have been substituted for the initial values.

After having been renamed it can thus serve as an input file in the next run. This is convenient if, e.g., the optimization algorithm has been stopped

---

[1] The reason is that the algorithm sometimes has to seek the minimum along a line, in order to obtain a sufficient reduction in F, in which case the iteration will involve several function evaluations. This is the "linear search subproblem" referred to in Section 3.2.

before having reached the optimum, due to a too low value having been assigned to MXFN. The search can then continue from the point where it was previously terminated.

Another conceivable use of the file is to provide starting values when the stochastic specification is changed from the one assuming time-independent residuals to the one allowing them to follow a first order vector autoregression.

APPENDIX A THE CONCENTRATED LOG-LIKELIHOOD FUNCTION AND ITS FIRST ORDER DERIVATIVES

The (unconcentrated) log-likelihood function corresponding to the equations (1), (2) and (3) is given by[1]

$$L(\Theta,\mathbf{B},\Sigma) = k + T\ell n|\mathbf{B}'| - \frac{1}{2}T\ell n|\Sigma| - \frac{1}{2}tr[(\mathbf{U}'\mathbf{U} - 2\mathbf{B}\mathbf{U}_1'\mathbf{U} + \mathbf{B}\mathbf{U}_1'\mathbf{U}_1\mathbf{B}')\Sigma^{-1}] \quad (a1)$$

where index 1 denotes one period lag and

$$k = -\frac{1}{2}nT\ell n(2\Pi).$$

It should be noticed that T is the number of observations on the model, which is one less the number of observations on the variables.

Application of matrix differentiation techniques,[2] yields the following first order conditions for $\mathbf{B}$ and Σ, respectively:

$$\mathbf{B}\mathbf{U}'\mathbf{U}_1 = \mathbf{U}'\mathbf{U}_1 \quad (a2)$$

and

$$T\Sigma^{-1} = \mathbf{U}'\mathbf{U} - \mathbf{U}'\mathbf{U}_1\mathbf{B}' - \mathbf{B}\mathbf{U}_1'\mathbf{U} + \mathbf{B}\mathbf{U}_1'\mathbf{U}_1\mathbf{B}' \quad (a3)$$

---

[1] Cf. Hendry (1971).

[2] Cf., e.g., Neudecker (1969) and Pollock (1979). Derivations based on Pollock's conventions can be found in Mellander (1984).

Since $\mathbf{H}$ and Σ are both unrestricted, they can be solved for from (a2) and (a3).[1] Substitution of the solutions into (a1) gives the concentrated log-likelihood function.

The solution for $\mathbf{H}$, implied by (a2), is

$$\begin{aligned}\tilde{\mathbf{H}} &= \mathbf{H}(\theta) \\ &= \mathbf{U}'\mathbf{U}_1(\mathbf{U}_1'\mathbf{U}_1)^{-1}.\end{aligned} \tag{a4}$$

By substitution of (a4) in (a3) Σ can also be expressed as a function of θ according to

$$\begin{aligned}\tilde{\Sigma} &= \Sigma(\theta) \\ &= T^{-1}(\mathbf{U}'\mathbf{U}) - \tilde{\mathbf{H}}\mathbf{U}'\mathbf{U}_1).\end{aligned} \tag{a5}$$

The concentrated log-likelihood function thus becomes

$$L^*(\theta) = k^* + T\ell n|\mathbf{B}| - \frac{1}{2}T\ell n|\tilde{\Sigma}| \ , \tag{a6}$$

where

$$k^* = -\frac{1}{2}nT[\ell n(2\Pi) + 1].$$

Differention of L* yields[2,3]

---

[1] The solutions for $\mathbf{H}$ and Σ, given below, can also be found in Hendry, op.cit.

[2] A derivation is given in Mellander (1984). Following Pollock (op.cit.) $\partial L^*/\partial\theta$ is defined to be a row vector; thus the transposition.

[3] The corresponding expression for the case when $\mathbf{A}$ is unrestricted, is given in Hendry (op.cit.). Due to a misprinting, the prime on $\tilde{\mathbf{H}}$ in the last term has been omitted there, however.

$$\left(\frac{\partial L^*}{\partial \boldsymbol{\theta}}\right)' = \left(\frac{\partial \mathrm{vec}\mathbf{A}}{\partial \boldsymbol{\theta}}\right)' \mathrm{vec}[\mathbf{T}(\mathbf{B}^{-1}:\mathbf{O}) - \mathbf{W}\mathbf{X} + \tilde{\mathbf{H}}'\mathbf{W}\mathbf{X}_1] \tag{a7}$$

where

$$\mathbf{W} = \tilde{\Sigma}^{-1}(\mathbf{U}' - \tilde{\mathbf{H}}\mathbf{U}_1'). \tag{a8}$$

The partitioned matrix $(\mathbf{B}^{-1}:\mathbf{O})$ has the same dimensions as the matrix **A**, and so the zero matrix has the same dimensions as the **C** matrix.

The n(n+m)xk matrix $\partial \mathrm{vec}\mathbf{A}/\partial\boldsymbol{\theta}$ can be written

$$\frac{\partial \mathrm{vec}\mathbf{A}}{\partial \boldsymbol{\theta}} = \left(\frac{\partial \mathrm{vec}\mathbf{A}}{\partial \theta_1}, \frac{\partial \mathrm{vec}\mathbf{A}}{\partial \theta_2}, \ldots, \frac{\partial \mathrm{vec}\mathbf{A}}{\partial \theta_k}\right). \tag{a9}$$

Denoting a particular element of vec**A** by a_g (cf. Section 2.2.2), the derivative of this coefficient with respect to the i:th parameter can be obtained according to

$$\frac{\partial a_g}{\partial \theta_i} = \frac{\partial a_g}{\partial h_\ell(\theta_i)} \frac{\partial h_\ell(\theta_i)}{\partial \theta_i} \tag{a10}$$

Expansion of the first factor on the RHS yields

$$\frac{\partial a_g}{\partial h_\ell(\theta_i)} = \left[d_g \cdot \frac{\partial n_g}{\partial h_\ell(\theta_i)} - n_g \cdot \frac{\partial d_g}{\partial h_\ell(\theta_i)}\right]/d_g^2 . \tag{a11}$$

where

$$\frac{\partial n_g}{\partial h_\ell(\theta_i)} = r_{gi} + 2\rho_{gij}h_\ell(\theta_i) + \sum_{j\neq i} \rho_{gij}h_\ell(\theta_j) \tag{a12}$$

and an analogous expression for $\partial d_g/\partial h_\ell(\theta_i)$, with r and ρ replaced by s and σ, respectively.

For $\ell=0$ the last factor on the RHS of (a10) equals unity (cf. Section 2.2.2). If $\ell=1$ then

$$\frac{\partial h_1(\theta_i)}{\partial \theta_i} = \begin{cases} [\theta_i/h_1(\theta_i)]^5 & \text{if } \gamma_i>0 \\ \theta_i/[h_1(\theta_i) + 0.05] & \text{if } \gamma_i=0 \end{cases} \quad \text{(a13)}$$

(cf. Section 2.2.1). The corresponding formula for $\ell=2$ is obtained by mere substitution of 2 and ν_i for 1 and γ_i, respectively, and multiplication by -1.

APPENDIX B AN EXAMPLE[1]

The following example is based on the disequilibrium version of the supply and demand model for exports, suggested by Goldstein and Khan (1978). The model will be estimated on Swedish annual data 1959-80, under the two alternative stochastic specifications that the program can allow for.

Overview of the model

Goldstein and Khan consider the following demand and supply equations for exports.

$$\log x_t^d = \alpha_0 + \alpha_1 \log(px/pxw)_t + \alpha_2 \log yw_t \qquad (b1)$$

where

x^d = quantity of exports demanded
px = price of exports
pxw = weighted average of the export prices of Sweden's trading partners
yw = weighted average of the real incomes of Sweden's trading partners

and

$$\log x_t^s = \beta_0 + \beta_1 \log(px/p)_t + \beta_2 y_t^* \qquad (b2)$$

where

---

[1] The dataset used in this example was kindly provided by Eva Christina Horwitz, Kommerskollegium, Stockholm.

x^s = quantity of exports supplied
p = domestic price index
y^* = logarithm of an index of domestic capacity.

As the equations have been specified in logarithmic form the parameters - except for the intercepts - can be regarded as elasticities. Accordingly, α_1 and β_1 are the price elasticities of export demand and export supply, respectively. Of course, α_1 is expected to be negative and β_1 to be positive. Further, the parameter α_2 can be interpreted as the aggregate income elasticity of the countries importing Swedish goods. This parameter should thus be positive (assuming Swedish exports not to be inferior goods). Finally, the capacity variable y^* is assumed to have a positive influence on the supply of exports.

The disequilibrium properties are introduced by the following equations[1]

$$\log x_t - \log x_{t-1} = \pi(\log x_t^d - \log x_{t-1}) \qquad \text{(b3)}$$

$$0 < \pi \leq 1$$

and

$$\log px_t - \log px_{t-1} = \lambda(\log x_t - \log x_t^s) \qquad \text{(b4)}$$

$$\lambda > 0$$

---

[1] In Goldstein's and Khan's notation the parameter in equation (b3) is denoted by γ instead of π. The change here has been made to avoid confusion with the constant γ_i used in the transformation (5a).

where x_t is the volume of exports actually realized in period t.[1] The first of these equations is a direct application of Nerlove's (1956) partial adjustment model. Its interpretation is that because of sluggishness in adjustment behavior, the actual change in the quantity of exports is only a fraction of the change necessary to eliminate all of the excess demand. The expected length of the time lag in the adjustment process can be estimated as $1/\pi$.[2]

It should be noted that the specification (b3) is not in accordance with the usual assumption of price-taking behavior on the part of the producers/exporters. For instance, if the demand for exports is increased, (b3) postulates an increase in the quantity of exports rather than a rise in the export price, which would have occurred had the exporters been price takers.

---

[1] The careful reader may notice that whereas the actual quantity of exports enters the RHS of (b3) with a one period lag, no such lag is present on RHS of (b4). This is for reasons of simplicity only. The dynamic counterparts of (b3) and (b4) are, respectively

$$\frac{d}{dt}\,\mathrm{logx}(t) = \pi[\mathrm{logx}^{d}(t)-\mathrm{logx}(t)] \qquad (*)$$

and

$$\frac{d}{dt}\,\mathrm{logpx}(t) = \lambda[\mathrm{logx}(t)-\mathrm{logx}^{s}(t)]. \qquad (**)$$

Since $\mathrm{logx}^{d}(t)$ is obviously dependent upon logpx(t) the equations (*) and (**) are simultaneously determined. The equations (b3) and (b4) are just two of the many discrete approximations to (*) and (**) which preserve this property. Had logx been lagged in (b4) the resulting system would not have had the simultaneity property - it would have been recursive instead of simultaneous.

[2] The reason is that the distribution of adjustment times corresponding to equation (*) is exponential with intensity π.

Obviously, if the exporters are not price takers they must be price setters. This is also the meaning of the second adjustment equation (b4), which says that the export price adjusts to conditions of excess supply. Accordingly, if the volume of exports that the exporters would like to supply at the going prices, x_t^s, is larger than the actual quantity, x_t, export prices will be lowered.

For a rather small country like Sweden, it might seem somewhat strange to assume exporters to be price setters. However, Goldstein and Khan obtained quite plausible results with this set-up, even for very small countries like Belgium and the Netherlands, pointing to the possibility that the supply side may have a more monopolistic structure than the demand side.

To obtain the final model, (b1) and (b2) are inserted in (b3) and (b4), respectively. This yields, after some rearranging, the following simultaneous equation system, containing observable variables only:

$$\log x_t = (\pi\alpha_0) + (\pi\alpha_1)\log(px/pxw)_t + (\pi\alpha_2)\log yw_t + (1-\pi)\log x_{t-1} \tag{b5}$$

$$\log px_t = -\frac{\lambda\beta_0}{1+\lambda\beta_1} + \frac{\lambda}{1+\lambda\beta_1}\log x_t + \frac{\lambda\beta_1}{1+\lambda\beta_1}\log p_t - \frac{\lambda\beta_2}{1+\lambda\beta_1} y_t^* + \frac{1}{1+\lambda\beta_1}\log px_{t-1} \cdot \tag{b6}$$

Although simple, this is a quite flexible, dynamic model. In contrast to the static relationships (b1) and (b2) which constituted the starting point, it allows the effects of changes in the exogenous variables to be dependent upon the time perspective considered. Both

the short-run and the long-run impacts can be calculated, as well as the effects after an arbitrary number of periods. However, although CONRAD generates all the output necessary to obtain these so called dynamic multipliers, the computational details lie outside the scope of this example. The interested reader is instead referred to Wallis (1973) and Brissimis (1976).

Estimation

To estimate the system (b5)-(b6) by means of CONRAD we first rename the parameters according to Table 1, which also gives the expected signs (if any) of the parameters.

Table 1 Parameter notation

Original notation	CONRAD notation	Expected sign
π	θ_1	+ (≤ 1)
α_0	θ_2	
α_1	θ_3	-
α_2	θ_4	+
λ	θ_5	+
β_0	θ_6	
β_1	θ_7	+
β_2	θ_8	+

Secondly, we form the matrices of endogenous and predetermined variables. They are

$$\mathbf{Y} = (\mathbf{y}_{.1}, \mathbf{y}_{.2}) = (\log \mathbf{x}, \log \mathbf{px}) \tag{b7a}$$

and[1]

$$\begin{aligned} \mathbf{Z} &= (\mathbf{z}_{.1}, \mathbf{z}_{.2}, \mathbf{z}_{.3}, \mathbf{z}_{.4}, \mathbf{z}_{.5}, \mathbf{z}_{.6}, \mathbf{z}_{.7}) \\ &= (\mathbf{1}, \log \mathbf{pxw}, \log \mathbf{yw}, \log \mathbf{p}, \mathbf{y}^*, \log \mathbf{x}_{-1}, \log \mathbf{px}_{-1}) \end{aligned} \tag{b7b}$$

respectively, where **1** is the unit vector.

The system (b5)-(b6) is thus reformulated according to

$$\log x_t = a_{12} \log px_t + a_{13} + a_{14} \log pxw_t + a_{15} \log yw_t + a_{18} \log x_{t-1} \tag{b8}$$

$$\log px_t = a_{21} \log x_t + a_{23} + a_{26} \log p_t + a_{27}\, y_t^* + a_{29} \log px_{t-1} \tag{b9}$$

where

$$a_{12} = \theta_1\theta_3 \,,\quad a_{13} = \theta_1\theta_2 \,,\quad a_{14} = -\theta_1\theta_3 \,,$$

$$a_{15} = \theta_1\theta_4 \,,\quad a_{18} = 1 - \theta_1 \,,$$

$$a_{21} = \frac{\theta_5}{1+\theta_5\theta_7} \,,\quad a_{23} = -\frac{\theta_5\theta_6}{1+\theta_5\theta_7} \,,$$

$$a_{26} = \frac{\theta_5\theta_7}{1+\theta_5\theta_7} \,,\quad a_{27} = -\frac{\theta_5\theta_8}{1+\theta_5\theta_7} \,,$$

---

[1] As explained in Section 2.2.2 $y_{.j}$ denotes the j:th column of the y matrix, and so forth.

$$a_{29} = \frac{1}{1+\theta_5\theta_7} \quad .$$

Regarding identification (cf. Section 2.4), it can be seen that problems are bound to occur whenever θ_1, and/or θ_5 are close to zero, since several of the columns of $\partial vec\mathbf{A}/\partial\boldsymbol{\theta}$ will then approach zero vectors. In fact, the latter parameter can cause trouble even without being particularly small, if the product $\theta_5\theta_7$ is large.

Actually, it is not necessary to inspect the matrix $\partial vec\mathbf{A}/\partial\boldsymbol{\theta}$ to see that these situations can cause trouble. According to (b3) and (b4) the quantity and price adjustment equations are not defined when $\theta_1=\pi$ and $\theta_5=\lambda$ are equal to zero. Further, if θ_7, and so $\theta_5\theta_7$, becomes very large all but one of the coefficients on the right hand side of (b9) will approach zero.

Whereas it is impossible to impose a constraint preventing the product $\theta_5\theta_7$ from becoming "too large", the hyperbolic transformation (5a) can be used to keep π and λ from taking on zero values. This can be done by defining

$$\pi \equiv h_1(\theta_1) \ , \qquad \gamma_1 > 0 \tag{b10}$$

and

$$\lambda \equiv h_1(\theta_5) \ , \qquad \gamma_5 > 0. \tag{b11}$$

To ensure that π and λ be strictly positive, without imposing too severe constraints, γ_1 and γ_5 can be assigned small positive values. If the restrictions are not binding the right hand sides of (b10) and (b11) will be very close to θ_1 and θ_5, respectively, making the differences between these definitions and the ones used in Table 1 very small.

Regarding the parameter π it should be noticed that while (b10) guarantees that $\pi>0$, there is nothing to stop π from taking on values larger than unity. Thus, one cannot be sure that $0<\pi\leq 1$ as required by (b3). Further, if it should happen that the constraint (b10) is not binding, but, instead, π takes on a value above unity it is not possible to impose the constraint $\pi\leq 1$, cf. Section 2.2.1.[1]

Concerning the stochastic specification, finally, the model (b8)-(b9) provides a good example of when the availability of the autoregressive specification (2) is of great value. The reason is that for models which, like this one, contain lagged endogenous variables, autocorrelation in the residuals will lead to inconsistent parameter estimates.[2] By means of the specification (2) it will be possible to test if the residuals are autocorrelated and, if this should be the case, to correct for it so as to obtain consistent estimates.

The estimations reported below were carried out on the DEC-10 computer at the Stockholm University Computing Center (QZ). Although not explicitly recorded here the same examples have been executed on the other machine on which CONRAD has been implemented - a PRIME 750 at the Stockholm School of Economics. As discussed in Section 3, the PRIME is inferior to the DEC-10 regarding the degree of precision in the calculations. Accordingly, the two computers yield results which differ slightly in some respects. These differences will be commented upon at appropriate places.

---

[1] An attempt to impose an erroneous constraint like $\pi\leq 1$ will cause the program to send a message to the terminal saying that the limit in one of the hyperbolic transformations has been incorrectly specified, whereupon the execution will be terminated.

[2] See, e.g., Maddala (1977, pp. 371-72).

Input

DATAM.DAT

```
0.59333  4.34251 1.  4.42004  4.03954  4.30136  4.08933  0.50682  4.34251
0.72271  4.36437 1.  4.44030  4.15779  4.32413  4.16200  0.59333  4.34251
0.81978  4.37450 1.  4.45085  4.19192  4.35927  4.22975  0.72271  4.36437
0.92028  4.37450 1.  4.44969  4.25731  4.36945  4.29456  0.81978  4.37450
1.00063  4.39445 1.  4.45783  4.32744  4.37450  4.35543  0.92028  4.37450
1.10856  4.40672 1.  4.47050  4.44524  4.41764  4.41159  1.00063  4.39445
1.18173  4.43082 1.  4.48526  4.54287  4.45899  4.46591  1.10856  4.40672
1.26976  4.45435 1.  4.50756  4.63337  4.47392  4.51743  1.18173  4.43082
1.31909  4.48864 1.  4.51305  4.69583  4.46706  4.56539  1.26976  4.45435
1.39377  4.48864 1.  4.50976  4.83961  4.45202  4.61115  1.31909  4.48864
1.53687  4.52179 1.  4.54436  4.96375  4.51415  4.65586  1.39377  4.48864
1.65250  4.60517 1.  4.60517  5.04349  4.60517  4.69866  1.53687  4.52179
1.70656  4.65396 1.  4.65110  5.11883  4.65205  4.73883  1.65250  4.60517
1.73519  4.78749 1.  4.73180  5.20351  4.77576  4.77828  1.70656  4.65396
1.88555  4.96284 1.  4.89485  5.33383  4.97949  4.81624  1.73519  4.78749
1.95445  5.15329 1.  5.09006  5.43149  5.18122  4.85203  1.88555  4.96284
1.88555  5.34711 1.  5.20483  5.38564  5.34568  4.88734  1.95445  5.15329
1.90658  5.38907 1.  5.21330  5.48935  5.36317  4.92071  1.88555  5.34711
1.89912  5.45104 1.  5.29079  5.54869  5.42495  4.95371  1.90658  5.38907
1.95303  5.53733 1.  5.44415  5.58953  5.51423  4.98498  1.89912  5.45104
2.01357  5.69709 1.  5.57405  5.63679  5.67332  5.01595  1.95303  5.53733
1.96991  5.84064 1.  5.67195  5.69235  5.82393  5.04600  2.01357  5.69709
```

The 22 observations cover the period 1959-80. However, these 22 observations on the variables will correspond to only 21 observations on the model, namely the years 1960-80. The reason is that one observation is lost when the stochastic specification (2) is used.

In the first estimation, when the residuals are assumed to be only contemporaneously correlated, i.e. when $\mathbf{H}$ is equal to the zero matrix, it would be possible to use all 22 observations. But, in order to make the LR test of the hypothesis $\mathbf{H}=\mathbf{0}$ applicable, the first observation will be disregarded under this more restrictive stochastic specification, too. (Cf. Section 2.5.)

PARIC.DAT

```
8
0.49 1 1 0.1
-2.73 2 0 /
-1.15 3 0 /
1.11 4 0 /
0.38 5 1 0.1
5.65 7 0 /
-4.97 6 0  /
1.77 8 0 /
```

The initial values have been obtained by 2SLS. Accordingly, logx and logpx have first been regressed on all the predetermined variables (the z_i's), yielding predicted values $\widehat{logx}$ and $\widehat{logpx}$, respectively. After substitution of the endogenous variables on the right hand sides of (b8) and (b9) for these predicted variables the equations have been estimated by OLS.[1] The resulting estimated equations where

$$logx_t = -0.563\ \widehat{logpx}_t - 1.340 + 0.444\ logpxw_t + 0.544\ logyw_t + 0.509\ logx_{t-1}$$

and

$$logpx_t = 0.127\ \widehat{logx}_t + 0.630 + 0.715\ logp_t - 0.224\ y_t^* + 0.336\ logpx_{-1}.$$

By means of the functional relations between the a_{ij}'s and the θ_i's given in connection with (b8) and (b9), the estimated coefficients yield the initial values in PARIC.DAT. For instance, an estimate of θ_1, is obtained as one minus the coefficient for $logx_{t-1}$. Given this estimate, initial estimates of θ_2, θ_3 and θ_4 are easily obtained from the other coefficients in the equation for $logx_t$. Likewise, an estimate of θ_5 can be obtained from the second equation by dividing the coefficient of $\widehat{logx}_t$ with the coefficient of $logpx_{t-1}$, whereupon θ_6, θ_7 and θ_8 can be solved for.

Notice that the initial values are not strictly ordered according to parameter index in PARIC.DAT; the initial value for θ_7 preceeds the initial value for θ_6. This is just to examplify the fact that the ordering of the initial values in this input file can be chosen freely

---

[1] In both the first and the second steps the estimations were based on data for the period 1960-80, i.e. on the same data set as the one subsequently employed by CONRAD.

by the user. As will be seen below, however, all output concerning the parameters is always ordered according to parameter index, starting with index 1 and ending with index k. The reason is that program automatically sorts the content of PARIC.DAT in this order.

It can also be seen that inequality constraints have been imposed on π and λ, in accordance with (b10) and (b11). For both π and λ the lower limits has been set equal to 0.1.

GICOF.DAT

```
0 1.E-9 0 78 1
2 7 22
1  /
2 22 2 7
LOGX    LOGPX   CONST.  LOGPXW  LOGYW   LOGP    YSTAR   LOGX-1  LOGPX-1
EXPORT SYSTEM ACC. TO GOLDSTEIN & KHAN (1978). ESTIMATION PERIOD 1960-80
12 1 2 /
1 1 0
-1.0 /
1 3 1
0.0 0 1  /
1.0 1 2
1 4 1
0.0 0 1 /
-1.0 1 3
1 5 1
0.0 0 1 /
1.0 1 4
1 8 1
1.0 1 0 /
-1.0 1
2 1 2
0.0 1 0 1.0 0 1
1.0 5
1.0 5 7
2 2 0
-1.0 /
2 3 2
0.0 0 1 1.0 0 1
-1.0 5 6
1.0 5 7
2 6 2
0.0 0 1 1.0 0 1
1.0 5 7
1.0 5 7
2 7 2
0.0 0 1 1.0 0 1
-1.0 5 8
1.0 5 7
2 9 2
1.0 0 0 1.0 0 1
1.0 5 7
1 2 1
0.0 0 1 /
1.0 1 3
```

Since the first estimation is to be performed with **H=0** LP, the first number in the first row, has been set equal to zero.

The numerical value assigned to EPS, 1.E-9, might seem remarkably low. According to Section 3.2 it implies that the algorithm should keep searching until the changes in the parameter estimates are less than one over a billion in absolute value. Such an extremely high degree of precision in the $\hat{\theta}_i$'s is, of course, never needed. However, in addition to determining the precision in the parameter estimates themselves, EPS also, indirectly, affects the estimation of their variances and covariances. According to our experience, to get good estimates of the $\hat{\theta}_i$'s variances and covariances, EPS should not be set larger than 1.E-6. Often the increase in execution time caused by even lower EPS-values is quite modest, however.[1] On the other hand, values below 1.E-9 mostly only result in "underflow", i.e. numbers too small to be recognized by the computer.

MXFN has been set equal to zero, in order to obtain a listing of the input specifications in RESF1.DAT, before the actual estimation is carried out. By the setting of WIDTH=78 the printout is certain to be readable on an 80-character terminal.

Notice that IFO=2 (first number on the fourth line). The first row of the matrix in DATAM.DAT should thus be ignored; cf. the discussion above.

Since the model contains only nine variables altogether (including the intercept) all the variable names can be

---

[1] For example, regarding the present problem the CPU-time was increased only by 7 % when EPS was decreased from 1.E-6 to 1.E-9, i.e. by a thousand times.

written on one line. As mentioned in Section 4.3 nine is the maximum number of names that can be given on one line. Accordingly, if yet another variable is incorporated in the model its name will have to be given on a subsequent line.

The specifications of the nonzero coefficients (the a_{ij}'s) are given on line 8 and onwards. It might be somewhat difficult to see where different specifications start and end. As will be seen below, the corresponding listing in RESF1.DAT is easily readable, however. By means of RESF1.DAT it is thus easy to check if any of the coefficients have been wrongly specified. Further, since, in contrast to the parameter specifications, the program prints out the coefficient specifications in the same order as they have been input it should not be very difficult to see where in GICOF.DAT changes should be made, if errors are discovered.

Results[1]

Execution of CONRAD by the command "RUN CONRAD" now results in the file RESF1.DAT which looks like the following.

---

[1] In this section the convention of denoting vectors and matrices by boldface type has been abandoned, in order to conform with the output produced by the program.

RESF1.DAT

EXPORT SYSTEM ACC. TO GOLDSTEIN & KHAN (1978). ESTIMATION PERIOD 1960-80

Y-MATRIX

LOGX	LOGPX
0.722710	4.364370
0.819780	4.374500
0.920280	4.374500
1.000630	4.394450
1.108560	4.406720
1.181730	4.430820
1.269760	4.454350
1.319090	4.488640
1.393770	4.488640
1.536870	4.521790
1.652500	4.605170
1.706560	4.653960
1.735190	4.787490
1.885550	4.962840
1.954450	5.153290
1.885550	5.347110
1.906580	5.389070
1.899120	5.451040
1.953030	5.537330
2.013570	5.697090
1.969910	5.840640

Z-MATRIX

CONST.	LOGPXW	LOGYW	LOGP	YSTAR	LOGX-1	LOGPX-1
1.000000	4.440300	4.157790	4.324230	4.162000	0.593330	4.342510
1.000000	4.450850	4.191920	4.359270	4.229750	0.722710	4.364370
1.000000	4.449690	4.257310	4.369450	4.294560	0.819780	4.374500
1.000000	4.457830	4.327440	4.374500	4.355430	0.920280	4.374500
1.000000	4.470500	4.445240	4.417640	4.411590	1.000630	4.394450
1.000000	4.485260	4.542870	4.458990	4.465910	1.108560	4.406720
1.000000	4.507560	4.633370	4.473920	4.517430	1.181730	4.430820
1.000000	4.513050	4.695830	4.467060	4.565390	1.269760	4.454350
1.000000	4.509760	4.839610	4.452020	4.611150	1.319090	4.488640
1.000000	4.544360	4.963750	4.514150	4.655860	1.393770	4.488640
1.000000	4.605170	5.043490	4.605170	4.698660	1.536870	4.521790
1.000000	4.651100	5.118830	4.652050	4.738830	1.652500	4.605170
1.000000	4.731800	5.203510	4.775760	4.778280	1.706560	4.653960
1.000000	4.894850	5.333830	4.979490	4.816240	1.735190	4.787490
1.000000	5.090060	5.431490	5.181220	4.852030	1.885550	4.962840
1.000000	5.204830	5.385640	5.345680	4.887340	1.954450	5.153290
1.000000	5.213300	5.489350	5.363170	4.920710	1.885550	5.347110
1.000000	5.290790	5.548690	5.424950	4.953710	1.906580	5.389070
1.000000	5.444150	5.589530	5.514230	4.984980	1.899120	5.451040
1.000000	5.574050	5.636790	5.673320	5.015950	1.953030	5.537330
1.000000	5.671950	5.692350	5.823930	5.046000	2.013570	5.697090

21 OBSERVATIONS; NUMBER 2 TO 22

PARAMETERS, INEQUALITY CONSTRAINTS

INIT. VALUES	INDEX	TRANSF.	BOUND
0.490006	1	1	0.10
-2.730000	2	0	
-1.150000	3	0	
1.110000	4	0	
0.380021	5	1	0.10
-4.970000	6	0	
5.650000	7	0	
1.770000	8	0	

REQUIRED ACCURACY IN THE ESTIMATES: .10E-08

EQUALITY CONSTRAINTS

A(1, 1) = -1.00000 TYPE OF CONSTRAINT: 0

R0 = -1.00

A(1, 3) = -1.33772 TYPE OF CONSTRAINT: 1

CONSTANTS GREEK R 1.00
CORR. PAR.-INDICES 1 2

A(1, 4) = 0.56351 TYPE OF CONSTRAINT: 1

CONSTANTS GREEK R -1.00
CORR. PAR.-INDICES 1 3

A(1, 5) = 0.54391 TYPE OF CONSTRAINT: 1

CONSTANTS GREEK R 1.00
CORR. PAR.-INDICES 1 4

A(1, 8) = 0.50999 TYPE OF CONSTRAINT: 1

R0 = 1.00

CONSTANTS R -1.00
CORR. PAR.-INDEX 1

A(2, 1) = 0.12075 TYPE OF CONSTRAINT: 2

CONSTANTS R 1.00
CORR. PAR.-INDEX 5

S0 = 1.00

CONSTANTS GREEK S 1.00
CORR. PAR.-INDICES 5 7

A(2, 2) = -1.00000 TYPE OF CONSTRAINT: 0

R0 = -1.00

A(2, 3) = 0.60014 TYPE OF CONSTRAINT: 2

CONSTANTS GREEK R -1.00
CORR. PAR.-INDICES 5 6

S0 = 1.00

CONSTANTS GREEK S 1.00
CORR. PAR.-INDICES 5 7

A(2, 6) = 0.68225 TYPE OF CONSTRAINT: 2

CONSTANTS GREEK R 1.00
CORR. PAR.-INDICES 5 7

S0 = 1.00

CONSTANTS GREEK S 1.00
CORR. PAR.-INDICES 5 7

A(2, 7) = -0.21373 TYPE OF CONSTRAINT: 2

CONSTANTS GREEK R -1.00
CORR. PAR.-INDICES 5 8

S0 = 1.00

CONSTANTS GREEK S 1.00
CORR. PAR.-INDICES 5 7

A(2, 9) = 0.31775 TYPE OF CONSTRAINT: 2

R0 = 1.00

S0 = 1.00

CONSTANTS GREEK S 1.00
CORR. PAR.-INDICES 5 7

A(1, 2) = -0.56351 TYPE OF CONSTRAINT: 1

CONSTANTS GREEK R 1.00
CORR. PAR.-INDICES 1 3

The first observation in DATAM.DAT is not printed out, since it will not be considered in the estimation. Notice that the initial values are now strictly ordered according to parameter index. This ordering of the parameters will be used in all output files.

As mentioned in Section 5.1, regarding the parameters which are subject to hyperbolical transformations, the hyperbolically transformed initial value have been substituted for the original initial values. Accordingly, the initial values for θ_1 and θ_5 have been replaced by, respectively,

$$(0.49^6 + 0.1^6)^{1/6} \quad \text{and} \quad (0.38^6 + 0.1^6)^{1/6} .$$

That the mappings from the θ_i's onto the a_{ij}'s have been correctly specified can be checked easily. For instance, we know that [cf. (b9)]

$$a_{27} = - \frac{\theta_5\theta_8}{1+\theta_5\theta_7}.$$

Given the initial values in this file a_{27} should thus be equal to

$$- \frac{0.380021 \cdot 1.77}{1+0.380021 \cdot 5.65} \approx - 0.21373 ,$$

which is in accordance with the value computed by the program; cf. the specification of a_{27} under the heading "EQUALITY CONTRAINTS".

The actual estimation was carried out by setting MXFN=100 and, again, giving the command RUN CONRAD. The run required 3.55 seconds of CPU time and yielded the following results. (The terminal output is not reproduced.)

RESF2.DAT

```
================================================================================
 EXPORT SYSTEM ACC. TO GOLDSTEIN & KHAN (1978). ESTIMATION PERIOD 1960-80
================================================================================

IEXIT= 1     COND= 0.02324

F= -0.1639077E+03    LNDET(B)=  0.7642503E-01    LNDET(SIGMA)= -0.1545741E+02

LNDET((Y-YBAR)'(Y-YBAR))=  0.1638678E+01

YBAR

   1.5160     4.8440

MINIMUM AT

  0.430083  -3.482521  -1.844085   1.030875   0.409474  -3.988291   7.544305

  1.129218

GRADIENT

-0.760E-08  0.110E-06  0.155E-07  0.622E-06  0.266E-07 -0.470E-07 -0.164E-08

-0.248E-06

STD. DEV.

  0.136357   0.599532   1.048350   0.133026   0.513633   2.341830  10.327559

  0.563581


HESIAN(-1)

  0.185933E-01   0.614245E-01   0.126428E+00  -0.142794E-01   0.475636E-01
  0.614245E-01   0.359438E+00   0.597994E+00  -0.796003E-01   0.221698E+00
  0.126428E+00   0.597994E+00   0.109904E+01  -0.134729E+00   0.409275E+00
 -0.142794E-01  -0.796003E-01  -0.134729E+00   0.176960E-01  -0.499618E-01
  0.475636E-01   0.221698E+00   0.409275E+00  -0.499618E-01   0.263818E+00
 -0.209366E+00  -0.932450E+00  -0.180332E+01   0.211325E+00  -0.115726E+01
 -0.864498E+00  -0.407633E+01  -0.750447E+01   0.917548E+00  -0.506909E+01
  0.507117E-01   0.227047E+00   0.436617E+00  -0.514092E-01   0.280815E+00


 -0.209366E+00  -0.864498E+00   0.507117E-01
 -0.932450E+00  -0.407633E+01   0.227047E+00
 -0.180332E+01  -0.750447E+01   0.436617E+00
  0.211325E+00   0.917548E+00  -0.514092E-01
 -0.115726E+01  -0.506909E+01   0.280815E+00
  0.548417E+01   0.221905E+02  -0.131894E+01
  0.221905E+02   0.106658E+03  -0.540428E+01
 -0.131894E+01  -0.540428E+01   0.317624E+00
```

A

-1.000000	-0.793131	-1.497813	0.793131	0.443373	0.000000	0.000000
0.100136	-1.000000	0.399373	0.000000	0.000000	0.755460	-0.113076

0.569906	0.000000
0.000000	0.244540

SIGMA

0.000898	-0.000260
-0.000260	0.000291

STRUCTURAL FORM EQ., NR.: 1

OBS.	Y	YP	RES.
2	0.72271	0.74401	-0.02130
3	0.81978	0.83321	-0.01343
4	0.92028	0.91660	0.00368
5	1.00063	0.99560	0.00503
6	1.10856	1.09394	0.01462
7	1.18173	1.19133	-0.00960
8	1.26976	1.27218	-0.00242
9	1.31909	1.32720	-0.00811
10	1.39377	1.41645	-0.02268
11	1.53687	1.51520	0.02167
12	1.65250	1.61421	0.03829
13	1.70656	1.71124	-0.00468
14	1.73519	1.73769	-0.00250
15	1.88555	1.80204	0.08351
16	1.95445	1.93480	0.01965
17	1.88555	1.89104	-0.00549
18	1.90658	1.87120	0.03538
19	1.89912	1.92180	-0.02268
20	1.95303	1.98885	-0.03582
21	2.01357	2.01685	-0.00328
22	1.96991	2.03978	-0.06987

(COS(P))**2	D-W
0.9948	1.4975

P=ANGLE BETWEEN (Y-YBAR) AND (YP-YPBAR)

STRUCTURAL FORM EQ., NR.: 2

OBS.	Y	YP	RES.
2	4.36437	4.32975	0.03462
3	4.37450	4.36370	0.01080
4	4.37450	4.37660	-0.00210
5	4.39445	4.38158	0.01287
6	4.40672	4.42351	-0.01679
7	4.43082	4.45893	-0.02811
8	4.45435	4.47909	-0.02474
9	4.48864	4.47918	0.00946
10	4.48864	4.47851	0.01013
11	4.52179	4.53472	-0.01293
12	4.60517	4.61832	-0.01315
13	4.65396	4.67500	-0.02104
14	4.78749	4.77880	0.00869
15	4.96284	4.97612	-0.01328
16	5.15329	5.17425	-0.02096
17	5.34711	5.33418	0.01293
18	5.38907	5.39312	-0.00405
19	5.45104	5.44558	0.00546
20	5.53733	5.53004	0.00729
21	5.69709	5.67389	0.02320
22	5.84064	5.81896	0.02168

(COS(P))**2	D-W
0.9989	1.1380

P=ANGLE BETWEEN (Y-YBAR) AND (YP-YPBAR)

-B(-1)C

-1.681056	0.734774	0.410751	-0.555092	0.083085	0.527973	-0.179682
0.231038	0.073578	0.041131	0.699875	-0.104756	0.052869	0.226548

OMEGA

0.001282	-0.000327
-0.000327	0.000213

REDUCED FORM EQ., NR.: 1

OBS.	Y	YP	RES.
2	0.72271	0.76788	-0.04517
3	0.81978	0.84015	-0.02037
4	0.92028	0.91533	0.00495
5	1.00063	1.00543	-0.00480
6	1.10856	1.08268	0.02588
7	1.18173	1.16997	0.01176
8	1.26976	1.25382	0.01594
9	1.31909	1.33355	-0.01446
10	1.39377	1.42223	-0.02846
11	1.53687	1.50730	0.02957
12	1.65250	1.60736	0.04514
13	1.70656	1.69544	0.01112
14	1.73519	1.74390	-0.00871
15	1.88555	1.79842	0.08713
16	1.95445	1.92084	0.03361
17	1.88555	1.90014	-0.01459
18	1.90658	1.87082	0.03576
19	1.89912	1.92415	-0.02503
20	1.95303	1.99157	-0.03854
21	2.01357	2.03365	-0.02008
22	1.96991	2.05056	-0.08065

(COS(P))**2	D-W
0.9926	1.2471

P=ANGLE BETWEEN (Y-YBAR) AND (YP-YPBAR)

REDUCED FORM EQ., NR.: 2

OBS.	Y	YP	RES.
2	4.36437	4.33427	0.03010
3	4.37450	4.36574	0.00876
4	4.37450	4.37610	-0.00160
5	4.39445	4.38206	0.01239
6	4.40672	4.42091	-0.01419
7	4.43082	4.45775	-0.02693
8	4.45435	4.47749	-0.02314
9	4.48864	4.48063	0.00801
10	4.48864	4.48136	0.00728
11	4.52179	4.53175	-0.00996
12	4.60517	4.61380	-0.00863
13	4.65396	4.67389	-0.01993
14	4.78749	4.77967	0.00782
15	4.96284	4.96740	-0.00456
16	5.15329	5.17089	-0.01760
17	5.34711	5.33564	0.01147
18	5.38907	5.38954	-0.00047
19	5.45104	5.44808	0.00296
20	5.53733	5.53390	0.00343
21	5.69709	5.67590	0.02119
22	5.84064	5.82704	0.01360

(COS(P))**2	D-W
0.9992	1.2325

P=ANGLE BETWEEN (Y-YBAR) AND (YP-YPBAR)

According to IEXIT and COND the run has been successfully completed. This is confirmed by the first order derivatives, given below the heading "GRADIENT", which are very close to zero.

When the same estimation was carried out on the PRIME 750 computer the requirement that COND<1 was not fulfilled. In fact, COND was far above unity. The primary reason for this was that due to the lower precision on the PRIME the algorithm did not manage to get the first order derivatives sufficiently close to zero; the smallest derivative (in absolute value) produced by the PRIME was of the same magnitude as the largest of the derivatives reported here. This, in turn, yielded a high value on COND.[1] However, that does not mean that the estimation was a failure. The parameter estimates were very close to those obtained on the DEC-10, the differences being confined to the fifth or sixth decimal points. Although the differences in the estimated standard deviations were higher, they were in no case larger than 5 % in relative magnitude.

By means of the information given on the second and third lines, Berndt's generalized R^2 can be computed according to

$$\tilde{R}^2 = 1-\exp[\text{LNDET(SIGMA)}-2\cdot\text{LNDET(B)} - \text{LNDET}\{(\text{Y-YBAR})'(\text{Y-YBAR})\}] \approx 0.99999997 ,$$

indicating an astonishingly good fit for the model as a whole. However, $\tilde{R}^2$ has a general tendency to lie very close to unity, comparing, as it does, the residual variance of the system actually estimated with that of a corresponding system, employing constants as the only

---

[1] Cf. Section 3.2.

explanatory variables.[1] As a consequence, $\tilde{R}^2$ may well be over 0.999 without the model fitting the data particularly well, especially if the model contains many equations.

When, as in this case, the estimation is based on a short time series and the variables are defined in levels rather than differences, a very good fit is often obtained, causing $\tilde{R}^2$ to take on values extremely close to one. [$\tilde{R}^2$'s of almost the same magnitude were reported by Berndt and Khaled (1979 p. 1235)].

The final estimates are given below "MINIMUM AT". They are ordered according to parameter index, beginning with $\hat{\theta}_1$, and ending with $\hat{\theta}_8$. Some of them differ substantially from the initial estimates. In particular, the price elasticities of demand and supply, $\hat{\theta}_3$ and $\hat{\theta}_7$ respectively, have changed considerably, and so has the elasticity with respect to production capacity, $\hat{\theta}_8$.

Further, it should be noticed that the inequality constraints have not been binding; both $\hat{\theta}_1$ $(=\hat{\pi})$ and $\hat{\theta}_5$ $(=\hat{\lambda})$ are far from the lower limit 0.1. Accordingly, their effects on the estimation can safely be assumed to have been negligible.

The variances and covariances of the parameter estimates can be found in the matrix denoted by HESIAN(-1), where "(-1)" denotes inverse. (For a further explanation of the naming of this matrix, cf. Section 3.3.) Because of the setting of WIDTH=78 it has been necessary to partition HESIAN(-1). Its sixth, seventh and

---

[1] Here, the model contains intercepts in both equations as indicated by the typical elements of YBAR (cf. Section 5.3). However, if it had lacked constants in some of the equations, the hypothetical model of comparison would not have had any constants in the corresponding equations, either.

eighth columns have been printed below the columns one, two and three, respectively. The standard errors have been given below "STD.DEV.", are equal to the square roots of the diagonal elements of HESIAN(-1). As can be seen, all the parameters in the first equation, i.e. $\hat{\theta}_1$ to $\hat{\theta}_4$, have small standard errors, while the opposite is true for the parameter estimates in the second equation.

HESIAN(-1) is followed by the coefficient matrix A and the variance-covariance matrix of the structural form residuals, SIGMA. The matrix A is partitioned in the same way as HESIAN(-1).

Actual and predicted values of the endogenous variables are next printed, together with the corresponding residuals. The 〔COS(P)〕**2", where "**2" denotes squared, are goodness-of-fit measures, explained in Section 5.3. "D-W" denotes the Durbin-Watson statistic.

Since, in this model, the equations contain lagged endogenous variables the Durbin-Watson statistics are biased towards acceptance of the null hypothesis of no autocorrelation 〔Maddala (1977, p. 372)〕. In spite of this bias they do not confirm the hypothesis.[1] This is a strong indication that the residuals, really, are autocorrelated.

After the 〔COS(P)〕**2 and the Durbin-Watson statistic of the last structural form equation - no. 2 in this case - information about the reduced form of the model is given. The matrix -B(-1)C contains the reduced form coefficients and OMEGA is the dispersion matrix of the reduced form of residuals.

---

[1] On both the 5 % and the 1 % significance level they fall in the so called "inconclusive" region, cf. Maddala (op.cit. p. 284).

RESF3.DAT

```
|=========================================================================|
| EXPORT SYSTEM ACC. TO GOLDSTEIN & KHAN (1978). ESTIMATION PERIOD 1960-80 |
|=========================================================================|
ENTRY TO VA09A
   0    1
-0.1017042E+03
 0.4900000E+00 -0.2730000E+01 -0.1150000E+01  0.1110000E+01  0.3800000E+00
-0.4970000E+01  0.5650000E+01  0.1770000E+01
 0.4703049E+02  0.1924761E+02 -0.6445542E+01  0.5715055E+02 -0.2710501E+02
-0.9689873E+00 -0.1303505E+01  0.6453632E+01
   5   13
-0.1530926E+03
 0.4993829E+00 -0.3539129E+01 -0.1116546E+01  0.1027611E+01  0.8636964E+00
-0.5031286E+01  0.6164022E+01  0.1362236E+01
-0.2045742E+02 -0.5386628E+02  0.1758788E+02 -0.1577648E+03  0.1196911E+02
-0.2881381E+02 -0.3022693E+00 -0.1548906E+03
  10   20
-0.1621864E+03
 0.4391762E+00 -0.3262737E+01 -0.1613577E+01  0.9835287E+00  0.9460301E+00
-0.4917398E+01  0.6174196E+01  0.1351201E+01
-0.1517398E+02 -0.4267438E+02 -0.3465197E+01 -0.2290187E+03  0.2485098E+01
 0.9581849E+01  0.3846265E+00  0.4581320E+02
  15   26
-0.1638115E+03
 0.4684736E+00 -0.3309839E+01 -0.1524524E+01  0.9920842E+00  0.6144172E+00
-0.4597627E+01  0.4727149E+01  0.1277229E+01
 0.1948798E+00  0.1203899E+01  0.3636853E+00  0.7290565E+01  0.3316656E+00
-0.8477188E+00 -0.7405970E-01 -0.3955837E+01
  20   31
-0.1638705E+03
 0.4356130E+00 -0.3424158E+01 -0.1757057E+01  0.1018277E+01  0.5206323E+00
-0.4391763E+01  0.5714685E+01  0.1226816E+01
-0.8350007E+00 -0.6176059E+00 -0.3397790E+00 -0.4884678E+01  0.2195213E+00
-0.3912416E+00 -0.2226160E-01 -0.1262589E+01
  25   36
-0.1639062E+03
 0.4329031E+00 -0.3468688E+01 -0.1818242E+01  0.1027769E+01  0.4233840E+00
-0.4052408E+01  0.7190997E+01  0.1145024E+01
-0.4033553E+00 -0.1221455E+01 -0.2524264E-01 -0.6155986E+01 -0.6713585E+00
 0.1022567E+01 -0.3437729E-02  0.4850337E+01
  30   41
-0.1639076E+03
 0.4312030E+00 -0.3478550E+01 -0.1836182E+01  0.1029954E+01  0.4133154E+00
-0.4009267E+01  0.7493270E+01  0.1134171E+01
 0.2732952E-01 -0.2166061E+00 -0.3661181E-01 -0.1253157E+01 -0.4981456E-01
 0.1090229E+00  0.3535902E-02  0.5727773E+00
  35   46
-0.1639077E+03
 0.4300919E+00 -0.3482483E+01 -0.1844010E+01  0.1030866E+01  0.4094975E+00
-0.3988379E+01  0.7543484E+01  0.1129240E+01
-0.5573053E-03 -0.8691489E-03  0.2855870E-03 -0.2952027E-02 -0.2763452E-03
-0.7260643E-03 -0.5047692E-04 -0.3729842E-02
  40   51
-0.1639077E+03
 0.4300831E+00 -0.3482521E+01 -0.1844085E+01  0.1030875E+01  0.4094735E+00
-0.3988291E+01  0.7544305E+01  0.1129218E+01
-0.3900872E-06  0.3968512E-06  0.2605563E-07  0.2146736E-05 -0.2923256E-06
 0.5229298E-06 -0.8463531E-08  0.2375873E-05
  43   57    1
-0.1639077E+03
 0.4300831E+00 -0.3482521E+01 -0.1844085E+01  0.1030875E+01  0.4094735E+00
-0.3988291E+01  0.7544305E+01  0.1129218E+01
-0.7603473E-08  0.1096530E-06  0.1549682E-07  0.6223280E-06  0.2658621E-07
-0.4695014E-07 -0.1637510E-08 -0.2476716E-06
```

The number of function evaluations thus amounted to 57, quite far from the set limit of 100.

As suggested in Section 5.4, the final output file, RESF4.DAT, will be used to provide starting values for the second estimation.

RESF4.DAT

```
8
 0.430083     1       1      0.10
-3.482521     2       0   /
-1.844085     3       0   /
 1.030875     4       0   /
 0.409474     5       1      0.10
-3.988291     6       0   /
 7.544305     7       0   /
 1.129218     8       0   /
```

Accordingly, this file will be renamed to PARIC.DAT. Further, LP, MXFN and IFO in GICOF.DAT are changed from 0, 100 and 2 to 1, 0 and 1, respectively.

Execution of CONRAD now yields a new RESF1.DAT file. To save space, the specifications of the equality constraints have not been reproduced.

RESF1.DAT (second estimation)

```
================================================================================
 EXPORT SYSTEM ACC. TO GOLDSTEIN & KHAN (1978). ESTIMATION PERIOD 1960-80
================================================================================
```

Y-MATRIX

LOGX	LOGPX
0.593330	4.342510
0.722710	4.364370
0.819780	4.374500
0.920280	4.374500
1.000630	4.394450
1.108560	4.406720
1.181730	4.430820
1.269760	4.454350
1.319090	4.488640
1.393770	4.488640
1.536870	4.521790
1.652500	4.605170
1.706560	4.653960
1.735190	4.787490
1.885550	4.962840
1.954450	5.153290
1.885550	5.347110
1.906580	5.389070
1.899120	5.451040
1.953030	5.537330
2.013570	5.697090
1.969910	5.840640

Z-MATRIX

CONST.	LOGPXW	LOGYW	LOGP	YSTAR	LOGX-1	LOGPX-1
1.000000	4.420040	4.039540	4.301360	4.089330	0.506820	4.342510
1.000000	4.440300	4.157790	4.324130	4.162000	0.593330	4.342510
1.000000	4.450850	4.191920	4.359270	4.229750	0.722710	4.364370
1.000000	4.449690	4.257310	4.369450	4.294560	0.819780	4.374500
1.000000	4.457830	4.327440	4.374500	4.355430	0.920280	4.374500
1.000000	4.470500	4.445240	4.417640	4.411590	1.000630	4.394450
1.000000	4.485260	4.542870	4.458990	4.465910	1.108560	4.406720
1.000000	4.507560	4.633370	4.473920	4.517430	1.181730	4.430820
1.000000	4.513050	4.695830	4.467060	4.565390	1.269760	4.454350
1.000000	4.509760	4.839610	4.452020	4.611150	1.319090	4.488640
1.000000	4.544360	4.963750	4.514150	4.655860	1.393770	4.488640
1.000000	4.605170	5.043490	4.605170	4.698660	1.536870	4.521790
1.000000	4.651100	5.118830	4.652050	4.738830	1.652500	4.605170
1.000000	4.731800	5.203510	4.775760	4.778280	1.706560	4.653960
1.000000	4.894850	5.333830	4.979490	4.816240	1.735190	4.787490
1.000000	5.090060	5.431490	5.181220	4.852030	1.885550	4.962840
1.000000	5.204830	5.385640	5.345680	4.887340	1.954450	5.153290
1.000000	5.213300	5.489350	5.363170	4.920710	1.885550	5.347110
1.000000	5.290790	5.548690	5.424950	4.953710	1.906580	5.389070
1.000000	5.444150	5.589530	5.514230	4.984980	1.899120	5.451040
1.000000	5.574050	5.636790	5.673320	5.015950	1.953030	5.537330
1.000000	5.671950	5.692350	5.823930	5.046000	2.013570	5.697090

22 OBSERVATIONS ON THE VARIABLES; NUMBER 1 TO 22,
CORRESPONDING TO 21 OBSERVATIONS ON THE AUTOREGRESSIVE MODEL

PARAMETERS, INEQUALITY CONSTRAINTS

INIT. VALUES	INDEX	TRANSF.	BOUND
0.430094	1	1	0.10
-3.482521	2	0	
-1.844085	3	0	
1.030875	4	0	
0.409488	5	1	0.10
-3.988291	6	0	
7.544305	7	0	
1.129218	8	0	

REQUIRED ACCURACY IN THE ESTIMATES: .10E-08

Notice the remark below the Z matrix. Regarding the small differences between the initial values for θ_1 and θ_5 here and in RESF4.DAT, see the comments above to the file RESF1.DAT generated in the first estimation.

The actual estimation required only 35 % more CPU time than the first, considerably less complicated one. Of course, to a large extent this can be explained by the use of good starting values. According to RESF3.DAT (not reproduced) the number of function evaluations was only 47, i.e. fewer than in the first estimation. However, that CONRAD computes the derivatives of the log-likelihood function analytically instead of approximating them numerically also keeps down the increase in execution time. With more parameters to estimate this advantage would become even more pronounced.

RESF2.DAT (second estimation)

```
================================================================================
 EXPORT SYSTEM ACC. TO GOLDSTEIN & KHAN (1978). ESTIMATION PERIOD 1960-80
================================================================================

IEXIT= 1     COND= 0.00338

F= -0.1711345E+03    LNDET(B)=  0.1601129E+00    LNDET(SIGMA)= -0.1597830E+02

LNDET((Y-YBAR)'(Y-YBAR))=  0.1638678E+01

YBAR

   1.5160     4.8440

MINIMUM AT

  0.425316  -3.006924  -1.408521   0.933795   1.356911  -4.591157   2.713114

  1.293701

GRADIENT

 0.691E-07  0.897E-07 -0.101E-07  0.391E-06  0.215E-08 -0.390E-07  0.542E-10

-0.160E-06

STD. DEV.

  0.101124    0.423575    0.465504    0.089844    0.550841    0.801980    1.129590

  0.170235


HESIAN(-1)

  0.102261E-01   0.212683E-01   0.398629E-01  -0.519384E-02  -0.244534E-01
  0.212683E-01   0.179416E+00   0.145183E+00  -0.378404E-01   0.312150E-01
  0.398629E-01   0.145183E+00   0.216694E+00  -0.330075E-01  -0.600220E-01
 -0.519384E-02  -0.378404E-01  -0.330075E-01   0.807187E-02  -0.406066E-02
 -0.244534E-01   0.312150E-01  -0.600220E-01  -0.406066E-02   0.303425E+00
  0.336090E-01   0.261204E+00   0.145956E+00  -0.547172E-01  -0.524644E-01
  0.105815E-01  -0.116237E+00  -0.116139E-01   0.223394E-01  -0.367713E+00
 -0.737732E-02  -0.544858E-01  -0.314483E-01   0.114841E-01   0.147412E-01


  0.336090E-01   0.105815E-01  -0.737732E-02
  0.261204E+00  -0.116237E+00  -0.544858E-01
  0.145956E+00  -0.116139E-01  -0.314483E-01
 -0.547172E-01   0.223394E-01   0.114841E-01
 -0.524644E-01  -0.367713E+00   0.147412E-01
  0.643171E+00   0.404799E-01  -0.136253E+00
  0.404799E-01   0.127597E+01  -0.146168E-01
 -0.136253E+00  -0.146168E-01   0.289800E-01

A

 -1.000000  -0.599084  -1.278929   0.599084   0.397169   0.000000   0.000000
  0.289848  -1.000000   1.330739   0.000000   0.000000   0.786391  -0.374977


  0.574672   0.000000
  0.000000   0.213609
```

H

```
  0.084911  -0.265410
 -0.461199   0.220157
```

HA

```
 -0.161839   0.214541  -0.461786   0.050869   0.033724  -0.208716   0.099523
  0.525011   0.056139   0.882812  -0.276297  -0.183174   0.173130  -0.082554


  0.048796  -0.056694
 -0.265038   0.047028
```

SIGMA

```
  0.000918  -0.000492
 -0.000492   0.000389
```

STRUCTURAL FORM EQ., NR.: 1

OBS.	Y	YP	RES.
2	0.72271	0.73621	-0.01350
3	0.81978	0.82890	-0.00912
4	0.92028	0.91733	0.00295
5	1.00063	1.00149	-0.00086
6	1.10856	1.09073	0.01783
7	1.18173	1.19678	-0.01505
8	1.26976	1.27531	-0.00555
9	1.31909	1.33425	-0.01516
10	1.39377	1.40681	-0.01304
11	1.53687	1.49915	0.03772
12	1.65250	1.61451	0.03799
13	1.70656	1.71384	-0.00728
14	1.73519	1.74517	-0.00998
15	1.88555	1.79706	0.08849
16	1.95445	1.94405	0.01040
17	1.88555	1.91629	-0.03074
18	1.90658	1.87904	0.02754
19	1.89912	1.92915	-0.03003
20	1.95303	1.97178	-0.01875
21	2.01357	2.00407	0.00950
22	1.96991	2.03329	-0.06338

(COS(P))**2	D-W
0.9947	1.9128

P=ANGLE BETWEEN (Y-YBAR) AND (YP-YPBAR)

STRUCTURAL FORM EQ., NR.: 2

OBS.	Y	YP	RES.
2	4.36437	4.35370	0.01067
3	4.37450	4.37183	0.00267
4	4.37450	4.37723	-0.00273
5	4.39445	4.36940	0.02505
6	4.40672	4.42039	-0.01367
7	4.43082	4.44161	-0.01079
8	4.45435	4.47225	-0.01790
9	4.48864	4.46332	0.02532
10	4.48864	4.47581	0.01283
11	4.52179	4.55344	-0.03165
12	4.60517	4.61761	-0.01244
13	4.65396	4.66305	-0.00909
14	4.78749	4.78156	0.00593
15	4.96284	5.00879	-0.04595
16	5.15329	5.16075	-0.00746
17	5.34711	5.32214	0.02497
18	5.38907	5.40323	-0.01416
19	5.45104	5.42653	0.02451
20	5.53733	5.54222	-0.00489
21	5.69709	5.69064	0.00645
22	5.84064	5.80832	0.03232

(COS(P))**2	D-W
0.9985	2.0761

P=ANGLE BETWEEN (Y-YBAR) AND (YP-YPBAR)

-B(-1)C

-1.768981	0.510448	0.338407	-0.401412	0.191406	0.489648	-0.109036
0.818003	0.147952	0.098087	0.670043	-0.319498	0.141924	0.182005

B(-1)HB

0.405885	-0.154143
-0.407366	-0.100817

B(-1)HC

0.844093	-0.184377	-0.122235	0.266210	-0.126937	-0.176864	0.072311
-0.638153	0.222855	0.147744	-0.095969	0.045761	0.213774	-0.026068

OMEGA

0.001195	-0.000271
-0.000271	0.000131

REDUCED FORM EQ., NR.: 1

OBS.	Y	YP	RES.
2	0.72271	0.73966	-0.01695
3	0.81978	0.82892	-0.00914
4	0.92028	0.91637	0.00391
5	1.00063	1.01415	-0.01352
6	1.10856	1.08639	0.02217
7	1.18173	1.18904	-0.00731
8	1.26976	1.26536	0.00440
9	1.31909	1.34493	-0.02584
10	1.39377	1.41143	-0.01766
11	1.53687	1.48857	0.04830
12	1.65250	1.61378	0.03872
13	1.70656	1.70812	-0.00156
14	1.73519	1.74672	-0.01153
15	1.88555	1.78669	0.09886
16	1.95445	1.94178	0.01267
17	1.88555	1.92449	-0.03894
18	1.90658	1.87589	0.03069
19	1.89912	1.93721	-0.03809
20	1.95303	1.96651	-0.01348
21	2.01357	2.00877	0.00480
22	1.96991	2.04041	-0.07050

(COS(P))**2	D-W
0.9931	1.9295

P=ANGLE BETWEEN (Y-YBAR) AND (YP-YPBAR)

REDUCED FORM EQ., NR.: 2

OBS.	Y	YP	RES.
2	4.36437	4.35861	0.00576
3	4.37450	4.37447	0.00003
4	4.37450	4.37610	-0.00160
5	4.39445	4.37332	0.02113
6	4.40672	4.41396	-0.00724
7	4.43082	4.44373	-0.01291
8	4.45435	4.47097	-0.01662
9	4.48864	4.47081	0.01783
10	4.48864	4.48093	0.00771
11	4.52179	4.53945	-0.01766
12	4.60517	4.60638	-0.00121
13	4.65396	4.66351	-0.00955
14	4.78749	4.78490	0.00259
15	4.96284	4.98014	-0.01730
16	5.15329	5.15707	-0.00378
17	5.34711	5.33343	0.01368
18	5.38907	5.39433	-0.00526
19	5.45104	5.43757	0.01347
20	5.53733	5.54613	-0.00880
21	5.69709	5.68925	0.00784
22	5.84064	5.82875	0.01189

(COS(P))**2	D-W
0.9995	2.2513

P=ANGLE BETWEEN (Y-YBAR) AND (YP-YPBAR)

EIGENVALUES OF THE H MATRIX

REAL PART	IMAGINARY
-0.203808E+00	0.000000E+00
0.508876E+00	0.000000E+00

As far as IEXIT, COND and the first order derivatives are concerned, the estimation looks very good, slightly better even than the first one. However, this time there is an additional characteristic of the solution which should be checked immediately, namely if the eigenvalues of the H matrix lie within the unit circle [cf. (4), Section 2.1]. That this is indeed the case can be seen from the very last lines on the printout.

For this problem, the results produced by the PRIME 750 computer were closer to the DEC-10 results given here, than they were in the first estimation. For instance, COND was equal to 1.1 and so came quite close to satisfy the requirement of being less than unity.

By means of the F-values in this and the previous esti-

mation the LR test statistic for testing the hypothesis of no autocorrelation is found to be

$$-2(163.9077 - 171.1345) \approx 14.45$$

Since the critical value at the 1 % significance level is 13.28 the hypothesis is decisively rejected. The Durbin-Watson statistics also look much better now.

As this file is organized in the same way as the one obtained in the first estimation the output should not need to be commented upon. Besides the H matrix the only new information concerns the coefficient matrix HA of the lagged variables and the corresponding reduced form matrices B(-1)HB and B(-1)HC.

The parameter estimates are all unchanged with respect to sign. Two of them, $\hat{\theta}_5 (=\hat{\lambda})$ and $\hat{\theta}_7 (=\hat{\beta}_1)$, have changed dramatically in magnitude, however.

From an economic viewpoint, the results seem quite plausible. The price elasticities of demand $(\hat{\theta}_3=\hat{\alpha}_1)$ and supply $(\hat{\theta}_7=\hat{\beta}_1)$ lie between the estimates obtained by Ettlin (1977) and Lundborg (1981). That the income elasticity of demand $(\hat{\theta}_4=\hat{\alpha}_2)$ and the elasticity of supply with respect to capacity $(\hat{\theta}_8=\hat{\beta}_2)$ both are close to one seems reasonable. Regarding the adjustment parameters, it might well be that there is a mean time lag of slightly less than 2 1/2 years $(1/\hat{\theta}_1)$ in the adjustment of the export volumes to changes in excess demand.

Of course, the fact that the estimates have sensible interpretations makes it all the more encouraging that they are all significantly different from zero. By means of the standard errors it can be seen that each of the t-ratios are at least 2.4 in absolute value.

REFERENCES

Barten, A.P., 1969, "Maximum Likelihood Estimation of a Complete System of Demand Equations", European Economic Review, 1, pp. 7-73.

Belsley, D.A., 1980. "On the Efficient Computation of the Nonlinear Full-Information Maximum-Likelihood Estimator", Journal of Econometrics, 14, pp. 203-225.

Berndt, E.R., Hall, B.H., Hall, R.E. and Hausman, J.H., 1974, "Estimation and Inference in Non-linear Structural Models", Annals of Economic and Social Measurement, 3, pp. 653-666.

Berndt, E.R. and Khaled, M.S., 1979, "Parameter Productivity Measurement and Choice Among Flexible Functional Forms", Journal of Political Economy, 81, pp. 1220-1245.

Berndt, E.R. and Savin, N.E., 1975, "Estimation and Hypothesis Testing in Singular Equation Systems with Autoregressive Disturbances", Econometrica, 43, pp. 937-58.

Berndt, E.R. and Savin, N.E., 1977, "Conflict Among Criteria for Testing Hypotheses in the Multivariate Linear Regression Model", Econometrica, 45, pp. 1263-1278.

Breusch, T.S., 1979, "Conflict Among Criteria for Testing Hypothesis: Extensions and Comments", Econometrica, 47, pp. 203-207.

Brissimis, S.N., 1976, "Multiplier Effects for Higher than First-Order Linear Dynamic Econometric Models", Econometrica, 44, pp. 593-597.

Brown, B.W., 1981, "Sample Size Requirements in Full Information Maximum Likelihood Estimation", International Economic Review, 22, pp. 443-460.

Calzolari, G. and Panattoni, L., 1983, "Hessian and Approximated Hessian Matrices in Maximum Likelihood Estimation: A Monte Carlo Study", paper presented at the European Meeting of the Econometric Society, Pisa, Italy.

Carter, R.A.L. and Nagar, A.L., 1977, "Coefficients of Correlation for Simultaneous Equation Systems", Journal of Econometrics, 6, pp. 39-50.

Durbin, J., 1957, "Testing for Serial Correlation in Systems of Simultaneous Regression Equations", Biometrika, 44, pp. 370-377.

Durbin J., 1970, "Testing for Serial Correlation in Least Squares Regression When Some of the Regressors are Lagged Dependent Variables", Econometrica, 38, pp. 410-421.

Durbin, J. and Watson, G.S., 1950, "Testing for Serial Correlation in Least Squares Regression - I", Biometrika, 37, pp. 409-428.

Durbin, J. and Watson, G.S., 1951, "Testing for Serial Correlation in Least Squares Regression - II", Biometrika, 38, pp. 159-178.

Engle, R.F., 1984, "Wald, Likelihood Ratio and Lagrange Multiplier Tests in Econometrics", in: Z. Griliches and M.D. Intrilligator (eds.) Handbook of Econometrics, Vol. II, Amsterdam, North-Holland.

Ettlin, F., 1777, "Arbetskostnader, priser och exportmarknadsandelar", SEB:s kvartalstidskrift, 1-2, s. 46-61.

Farebrother, R.W., 1980, "The Durbin-Watson Test for Serial Correlation When There is No Intercept in the Regression", Econometrica, 48, pp. 1553-1563.

Fischer, F.M., 1966, The Identification Problem in Economics, New York, McGraw-Hill.

Fletcher, R., 1970, "A New Approach to Variable Metric Algorithms", Computer Journal, 13, pp. 317-322.

Fletcher, R. and Powell, M.J.D., 1963, "A Rapidly Convergent Descent Method for Minimization", Computer Journal, pp. 163-168.

Fomby, T.B., Hill, R.C. and Johnson, S.R., 1984, Advanced Econometric Methods, New York, Springer-Verlag.

Goldstein, M. and Khan, M.S., 1978, "The Supply and Demand for Exports: A Simultaneous Approach", Review of Economics and Statistics, LX, pp. 275-286.

Gouriéroux, C., Holly, A. and Monfort, A., 1982, "Likelihood Ratio Test, Wald Test and Kuhn-Tucker Test in Linear Models with Inequality Constraints on the Regression Parameter", Econometrica, 50, pp. 63-80.

Haessel, W., 1978, "Measuring Goodness of Fit in Linear and Nonlinear Models", Southern Economic Journal, 44, pp. 648-652.

Hausman, J.A., 1983, "Specification and Estimation of Simultaneous Equations Models", in: Z. Griliches and M.D. Intrilligator (eds.), Handbook of Econometrics, Vol I, Amsterdam, North-Holland.

Hendry, D.F., 1971, "Maximum Likelihood Estimation of Systems of Simultaneous Regression Equations with Errors Generated by a Vector Auto-regressive Process", International Economic Review, 12, pp. 257-271.

Hendry, D.F., 1974, Maximum Likelihood Estimation of Systems of Simultaneous Regression Equations with Errors Generated by a Vector Auto-regressive Process: A Correction", International Economic Review, 15, p. 260.

Hendry, D.F., 1976, "The Structure of Simultaneous Equations Estimators", Journal of Econometrics, 4, pp. 51-88.

Jansson, L. and Mellander, E., 1983, "Optimization under Non-linear Constraints", Economic Letters, 11, pp. 343-346.

Judge, G.G., Griffiths, W.E. Hill, R.C. and Lee, T.C., 1980, The Theory and Practice of Econometrics, New York, John Wiley and Sons.

Klein, L.R., 1953, A Textbook of Econometrics, Evanston, Ill.; Row, Peterson.

Lau, L.J., 1978, "Testing and Imposing Monotonicity, Convexity and Quasi-Convexity Constraints", in M. Fuss and D. McFadden (eds.) Production Economics: A Dual Approach to Theory and Applications, Vol. I, Amsterdam, North-Holland.

Lundborg, P., 1981, "Elasticities of Supply and Demand for Swedish Exports", Scandinavian Journal of Economics, 83, pp. 444-448.

Maddala, G.S., 1978, Econometrics, New York, McGraw-Hill.

Magnus, J.R. and Neudecker, H., 1979, "The Commutation Matrix: Some Properties and Statistical Applications", Annals of Statistics, 7, pp. 381-394.

Mellander, E., 1984, "Härledning av informationsmatrisen när residualerna genereras av en vektorautoregressiv process och parametrarna lyder under icke-lineära restriktioner" (Derivation of the information matrix when the errors are generated by a vector autoregressive process and the parameters obey non-linear contraints), mimeographed.

Nerlove, M., 1956, "Estimates of the Elasticities of Supply of Selected Agricultural Commodities", Journal of Farm Economics, 38, pp. 496-509.

Neudecker, H., 1969, "Some Theorems on Matrix Differentiation with Special Reference to Kronecker Products", Journal of the American Statistical Association, 64, pp. 953-963.

Pollock, D.S.G., 1979, The Algebra of Econometrics, Chichester, John Wiley and Sons.

Quandt, R.E., 1983, "Computational Problems and Methods", in: Z. Griliches and M.D. Intrilligator (eds.), Handbook of Econometrics, Vol. I, Amsterdam, North-Holland.

Rothenberg, T.J., 1971, "Identification in Parametric Models", Econometrica, 39, pp. 577-591.

Silvey, S.D., 1975, "Statistical Inference, London, Chapman and Hall.

Stone, J.R.N., 1954, "Linear Expenditure Systems and Demand Analysis: An Application to the Pattern of British Demand", Economic Journal, 64, pp. 511-27.

Theil, H., 1983, "Linear Algebra and Matrix Methods in Econometrics", in: Z. Griliches and M.D. Intrilligator (eds.), Handbook of Econometrics, Vol. I, Amsterdam, North-Holland.

Wallis, K.F., 1973, Topics in Applied Econometrics, London, Gray Mills.

Wonnacott, R.J. and Wonnacott, T.H., 1970, Econometrics, New York, John Wiley and Sons.

Zellner, A., 1962, "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias", Journal of the American Statistical Association, 57, pp. 348-368.