Large Sample Estimation and Testing Procedures
for Dynamic Equation Systems

by

Franz Palm and Arnold Zellner

Research Memorandum nr. 1978-10

Large Sample Estimation and Testing Procedures for Dynamic Equation Systems

by

Franz Palm and Arnold Zellner*

May 1978
Comments welcome

I. Introduction

In this paper we consider large sample estimation and testing procedures for parameters of dynamic equation systems with moving average error terms that are frequently encountered in econometric work, see e.g. Quenouille (1957) and Zellner and Palm (1974). As pointed out in Zellner and Palm (1974), three equation systems that are particularly relevant in econometric model-building are (1) the final equations (FEs), (2) the transfer functions (TFs), and (3) the structural equations (SEs). In the present work, we specify these equation systems and develop large sample "joint" or "system" estimation and testing procedures for each system of equations. These "joint" or "system" estimation procedures are iterative. They provide asymptotically efficient estimates of the parameters at the second step of iteration. The maximum likelihood estimator is obtained by iterating until convergence. The "joint" estimation methods provide parameter estimates that are more precise in large samples than those provided by single-equation procedures and the "joint" testing procedures are more powerful in large samples than those based on single-equation methods.

The aim of the paper is to present a unified approach for estimating and testing FE, TF and dynamic SE systems. In the paper we use the results of previous work on the asymptotic properties of the maximum likelihood (ML) estimator of the parameters of a dynamic model. We extend the recent work on efficient two-step estimation of dynamic models (e.g. Dhrymes and Taylor (1976), Hatanaka (1976), Palm (1977b), Reinsel (1976,1977)).

* Research financed by National Science Foundation Grants GS 40033 and SOC 7305547, income from the H.G.B. Alexander Endowment Fund, Graduate School of Business, U. of Chicago, and the Belgian National Science Foundation. The present paper is a revision of an earlier draft completed in 1974. The first author is presently at the Free University of Amsterdam.

Previous work related to present work includes that of Deistler (1975, 1976), Hannan (1969, 1971) and Hatanaka (1975), who have considered the identification problem for dynamic SE systems with moving average error terms. Maximum likelihood estimation of dynamic SEMs with moving average errors has been considered by Byron (1973), Phillips (1966) and Wall (1976) in the time domain, and for dynamic SEMs with stationary errors by Espasa and Sargan (1975) in the frequency domain. Spectral estimation methods for static SEMs with stationary errors have been proposed by Hannan and Terrell (1973) and by Espasa (1975). Among many other workers, Akaike (1973), Anderson (1975), Box and Jenkins (1970), Durbin (1959), Hannan (1975), Kang (1973), Maddala (1971), Nelson (1976), Nicholls (1976), Osborn (1976), Pesaran (1973), Pierce (1972), Reinsel (1976), and Wilson (1973) have considered estimation of parameters of single-equation or multi-equation ARMA and transfer function models. The problem of TF estimation in a single-equation context has been extensively studied in the "distributed lag" area. Closely related to our approach for FEs is the work of Nelson (1976) who considered joint estimation of a special FE system with diagonal MA matrices.

For a system of TFs, Wilson (1973) proposes an iterative procedure leading to a ML estimator. With respect to ML methods for TFs (e.g. Wilson (1973)) and dynamic SEMs (e.g. Byron (1973), Phillips (1966) and Wall (1976)), our approach is computationally more convenient to implement while having similar asymptotic properties. Many of the spectral methods apply to more general models, in the sense that the authors assume a stationary error process. For an extensive review of the literature, the reader is referred to Aigner (1971), Nicholls, Pagan and Terrell (1975) and Åström and Bohlin (1966). Finally, estimation methods for dynamic models with autoregressive errors, which have a long tradition in econometrics, are reviewed by Hendry (1976).

In what follows we shall specify the FE system that we consider in Section II and then go on to develop estimation and testing procedures for parameters of the FE system. In Section III, a TF system is specified and inference procedures for it are developed, while in Section IV the SE system is presented and procedures for analyzing it are developed. Section V is devoted to summary and discussion of the results with particular emphasis on relating them to the structure of econometric estimation procedures and on pointing to problems that remain to be analyzed.

II. Specification of and Estimation and Testing Procedures for Final Equations

Let $\underline{z}_t' = (z_{1t}, z_{2t},\ldots,z_{pt})$ be a vector of observable random variables generated by the following multivariate autoregressive-moving average (ARMA) process such as studied by Quenouille (1957):

$$\underset{p\times p}{H(L)}\ \underset{p\times 1}{\underline{z}_t} = \underset{p\times 1}{\bar{\underline{c}}_0} + \underset{p\times p}{F(L)}\ \underset{p\times 1}{\underline{e}_t} \qquad t=1,2,\ldots,T \tag{2.1.}$$

where $\bar{\underline{c}}_0' = (\bar{c}_1,\bar{c}_2,\ldots,\bar{c}_p)$ is a vector of constants, L is a lag operator such that $L^n\underline{z}_t \equiv \underline{z}_{t-n}$, $H(L) = \{h_{ij}(L)\}$ and $F(L) = \{f_{ij}(L)\}$ are pxp matrix lag operators with typical elements being finite degree polynomials in L, namely $h_{ij}(L)$ and $f_{ij}(L)$, respectively, and $\underline{e}_t$ is a px1 random error vector. We assume that $\underline{e}_t$ is normally distributed with

$$E\underline{e}_t = \underline{0} \quad \text{and} \quad E\underline{e}_t\underline{e}_{t'}' = \delta_{tt'}I_p \tag{2.2.}$$

for all t and t' where $\delta_{tt'}$ is the Kronecker delta. Note that contemporaneous and serial correlation as well as different variances for the error process in (2.2.) can be introduced through appropriate specification of F(L). We further assume that the inverse of H(L), $H^{-1}(L) = H^*(L)/|H(L)|$, exists, where $H^*(L)$ is the adjoint matrix associated with H(L) and $|H(L)|$ is the determinant of H(L) that is a scalar polynomial of finite degree in L with roots lying outside the unit circle.

The "final equations" (FEs) associated with (2.1.), obtained by multiplying both sides of (2.1.) on the left by $H^*(L)$, are given by:

$$|H(L)|\underline{z}_t = \underline{c}_0 + H^*(L)F(L)\underline{e}_t \tag{2.3a.}$$

or

$$\theta(L)\underline{z}_t = \underline{c} + A(L)\underline{e}_t \tag{2.3b.}$$

where $\underline{c}_0 = H^*(L)\ \bar{\underline{c}}_0$ and $\underline{c}' = d^{-1}\underline{c}'_0 = d^{-1}(c_1c_2,\ldots,c_p)$ are vectors of constants, $\theta(L) = |H(L)|/d$ and $A(L) = H^*(L)F(L)/d$, with d being a normalizing constant. In order to identify the system (2.3b.), we assume among other things that the roots of $|A(L)|$ are outside the unit circle, and that both sides of (2.3b.) do not have common factors. As pointed out in previous work, Zellner and Palm (1974, 1975), the AR polynomial $\theta(L)$ operates on each element of $\underline{z}_t$. Unless there is cancelling, the AR parts of the equations in (2.3.) should be of identical order and have the same parameters. Since it is often of interest to test that the AR parameters are the same in different equations and also for greater generality, we shall take up the problem of estimating parameters of the following system:

$$\theta_i(L)z_{it} = c_i + \underline{\alpha}_i'\underline{e}_t, \qquad i=1,2,\ldots,p, \tag{2.4.}$$

where $\theta_i(L) = 1 - \theta_{i1}L - \theta_{i2}L^2 - \ldots - \theta_{in_i}L^{n_i}$, with n_i given, $i=1,2,\ldots,p$, and $\underline{\alpha}_i'$ is the ith row of $A(L)$.

In connection with convenient estimation of the parameters in (2.4.), we express the error vector $A(L)\underline{e}_t$ in (2.3b.) as

$$\begin{aligned} A(L)\underline{e}_t &= A_0\underline{e}_t + A_1\underline{e}_{t-1} + \ldots + A_m\underline{e}_{t-m} \\ &= \underline{v}_t + G_1\underline{v}_{t-1} + \ldots + G_m\underline{v}_{t-m} \end{aligned} \tag{2.5.}$$

where $G_i = A_iA_0^{-1}$, $i=1,2,\ldots,m$, A_0 is assumed to be non-singular, and $\underline{v}_t=A_0\underline{e}_t$, which satisfies $E\underline{v}_t = \underline{0}$ and

$$E\underline{v}_t\underline{v}_t' = A_0A_0' \equiv \Omega_v \quad \text{and } E\underline{v}_t\underline{v}_{t'}'=0,\ t\neq t'. \tag{2.6.}$$

A typical element of $\underline{\varepsilon}_t=A(L)\underline{e}_t$, say the ith, ε_{it} may be represented as a moving average in one random variable (see e.g. Ansley and al. (1977), Palm (1977a) or Granger and Morris (1976)):

$$\varepsilon_{it} = \underline{\alpha}_i'\underline{e}_t = v_{it} + \lambda_{i1}v_{it-1} + \ldots + \lambda_{im}v_{it-m} \tag{2.7.}$$

where the λ_{ij}'s are such that they reproduce the autocorrelation structure of ε_{it}, i.e.

$$\omega_{ii}\sum_{h=0}^{m-j}\lambda_{ih+j}\lambda_{ih} = \sum_{h=0}^{m-j}\underline{a}'_{ih+j}\underline{a}_{ih}, \qquad j=0,1,\ldots,m,$$

with $\lambda_{i0}=1$, ω_{ii}, the ixi element of Ω_v, defined in (2.6.), and $\underline{a}_{ih}'$ being the ith row of A_h in (2.5.). Note that the v_{it}'s on the r.h.s. of (2.7.) are normally and independently distributed, each with zero mean and common variance ω_{ii}.

Each FE may be estimated separately using a single-equation nonlinear least squares or single-equation ML procedure. Joint estimation of the parameters in the system shown in (2.4.) will now be considered. We write the system of FEs as in Palm (1977b)

$$(2.8.)\quad \begin{bmatrix} \theta_1(L) & 0 & \cdots & 0 \\ 0 & \theta_2(L) & & \cdot \\ \cdot & & \ddots & \cdot \\ \cdot & & & 0 \\ 0 & \cdots & 0 & \theta_p(L) \end{bmatrix} \begin{bmatrix} z_{1t} \\ z_{2t} \\ \cdot \\ \cdot \\ z_{pt} \end{bmatrix} = \underline{c} + \underline{v}_t + \sum_{h=1}^{m} G_h \underline{v}_{t-h}$$

or alternatively as

$$(2.9.)\quad \underline{z}_t = W_{0t}\,\underline{c} + W_{1t}\,\underline{\theta} + \underline{u}_t$$

with $W_{0t} = I_p$

$$\underset{p\times k}{W_{1t}} = \begin{bmatrix} z_{1t-1} & z_{1t-2} & \cdots & z_{1t-n_1} & 0 & \cdots & & & & 0 \\ 0 & \cdots & & 0 & z_{2t-1} & z_{2t-2} & \cdots & z_{2t-n_2} & 0 \cdots & 0 \\ \cdot & & & & & & & & & \cdot \\ \cdot & & & & & & & & & \cdot \\ \cdot & & & & & & & & & \cdot \\ 0 & \cdots & & & & & & & & z_{pt-n_p} \end{bmatrix}$$

$$\underset{1\times k}{\underline{\theta}'} = (\theta_{11}, \theta_{12}, \theta_{13} \cdots\cdots \theta_{1n_1}, \theta_{21} \cdots\cdots \theta_{2n_2}, \cdots \theta_{pn_p})$$

$$k = \sum_{i=1}^{p} n_i$$

$$\underline{u}_t = \underline{v}_t + \sum_{h=1}^{m} G_h \underline{v}_{t-h}$$

and for a sample of T observations

$$(2.10.)\quad \underset{Tp\times 1}{\underline{z}} = \underset{(Tp\times p)}{W_0}\ \underset{(p\times 1)}{\underline{c}} + \underset{(Tp\times k)}{W_1}\ \underset{(k\times 1)}{\underline{\theta}} + \underset{(Tp\times 1)}{\underline{u}}$$

where $\underline{z}' = (\underline{z}_1', \underline{z}_2', \ldots, \underline{z}_T')$, $W_0' = (I_p, I_p, \ldots, I_p)$, $W_1' = (W_{11}', W_{12}', \ldots, W_{1T}')$,

and $\underline{u}' = (\underline{u}_1', \underline{u}_2', \ldots, \underline{u}_T')$]

Assuming initial conditions to be zero*, the vector $\underline{u}$ may be expressed in terms of $\underline{v}$

$\underline{u} = M\underline{v}$, where

$$M = \begin{bmatrix} I_p & 0 & \cdot & & & & 0 \\ G_1 & I_p & 0 & \cdot & & & \cdot \\ G_2 & G_1 & I_p & 0 & \cdot & & \cdot \\ \cdot & & & & & & \cdot \\ \cdot & & & & & & \cdot \\ G_m & G_{m-1} & \cdot & \cdot & I_p\ 0 \ldots & & \cdot \\ 0 & G_m & \cdot & \cdot & \cdot & \ldots & \cdot \\ & & \ddots & & & & \\ 0 & \cdot & \cdot & \cdot & 0 & G_m \ldots & I_p \end{bmatrix}$$

and $\underline{v}' = (\underline{v}_1', \underline{v}_2', \ldots, \underline{v}_T')$.
Under zero initial conditions, the likelihood function is

$$(2.11.)\quad L(\underline{\theta}, \underline{c}, M, \Omega_v, \underline{z}) \propto$$

$$|\Omega_v|^{-T/2} \exp\left[-\tfrac{1}{2}(\underline{z} - W_0\underline{c} - W_1\underline{\theta})' M'^{-1}(I_T \otimes \Omega_v^{-1}) M^{-1} (\underline{z} - W_0\underline{c} - W_1\underline{\theta})\right].$$

As shown in Palm (1977b), the first order conditions for a maximum of the log-likelihood function are

$$(2.12.)\quad \frac{\partial S}{\partial \underline{\beta}} = W'M'^{-1}(I_T \otimes \Omega_v^{-1}) M^{-1}\underline{u} = \underline{0}$$

where $S = -\frac{1}{2}(\underline{z} - W_0\underline{c} - W_1\underline{\theta})' M'^{-1}(I_T \otimes \Omega_v^{-1}) M^{-1}(\underline{z} - W_0\underline{c} - W_1\underline{\theta})$

$\underline{\beta}' = (\underline{c}', \underline{\theta}', \underline{\gamma}')$

$\underline{\gamma}' = \text{vec}\,[G_1\ \ G_2\ \ \ldots\ \ G_m]'$ with "vec" denoting the operation of vectorizing a matrix, stacking column after column,

* This assumption is basically made for the purpose of simplicity. One can also "backforecast" the values of the initial conditions for a set of FEs, as Box and Jenkins (1970) do for single equation ARMA models, or treat the initial conditions as unknown parameters (e.g. Phillips (1966)). Whether backforecasting improves the properties of estimators under all conditions is not known.

$W = [W_0 \quad W_1 \quad W_2]$

$W_2' = [\underline{v}_{ij}^{\ell'}]$, where $i=1,\dots,p$, $\ell=1,\dots,m$, $j=1,\dots,p$, and with typical row

$$\underline{v}_{ij}^{\ell'} = [\underbrace{0,0,\dots,0}_{\ell\ p \text{ times}} \quad \underbrace{0,\dots\ 0\ v_{j1}\ 0\dots 0}_{p \text{ elements}} \quad \underbrace{0\dots 0\ v_{j2}\ 0\dots 0}, \underbrace{0\dots 0\ v_{jT-\ell}\ 0\dots 0}]$$

ith position

For given Ω_v, the set of equations in (2.12.) is non-linear in the parameters of M. The solution of (2.12.) requires an iterative procedure.

An alternative to the exact ML solution of (2.12.) consists in approximating the first order conditions (2.12.). Using a lemma given by Dhrymes and Taylor* (1976), a two-step estimator of $\underline{\beta}$ with the same asymptotic properties as the ML estimator is

(2.13.) $$\hat{\hat{\underline{\beta}}} = \hat{\underline{\beta}} - \Gamma^{-1}(\hat{\underline{\beta}}) \frac{\partial S}{\partial \underline{\beta}} (\hat{\underline{\beta}})$$

where $\Gamma(\hat{\underline{\beta}})$ is a non-singular matrix such that

$$\operatorname*{plim}_{T \to \infty} \frac{1}{T} \Gamma (\hat{\underline{\beta}}) = \operatorname*{plim}_{T \to \infty} \frac{1}{T} \frac{\partial^2 S}{\partial \underline{\beta} \partial \underline{\beta}'} (\underline{\beta}_0) ,$$

$\underline{\beta}_0$ is the true parameter value and $\hat{\underline{\beta}}$ is a consistent estimator of $\underline{\beta}_0$ such that $\sqrt{T}\ (\hat{\underline{\beta}} - \underline{\beta}_0)$ has some limiting distribution. It should be noticed that the matrix Γ and the vector $\partial S/\partial \underline{\beta}$ in (2.13) depend on the unknown parameters of Ω_v . As the information matrix is block diagonal for $\underline{\beta}$ and Ω_v, the use of a consistent estimator ** for the unknown Ω_v in (2.12) will not affect the asymptotic properties of the solution of (2.12).

Application of generalized least squares (see e.g. Palm (1977b)) to

(2.14.) $$\underline{y} = W_0\ \underline{c} + W_1\ \underline{\theta} + W_2\ \underline{\gamma} + M\underline{v},$$
$$\underline{y} = W\underline{\beta} + \underline{u}$$

* For an earlier discussion on approximations to the ML solution, the reader is referred to Fisher (1925), chap. 9, Kendall and Stuart (1961), pp. 48-51, and Rothenberg and Leenders (1964).

** The lemma given by Dhrymes and Taylor (1976) applies to all the parameters in the likelihood function, i.e. to $\underline{\theta}' = (\underline{\beta}'\ ,\ \text{vec}\ (\Omega_v)')$. Using a block-diagonal matrix Γ to approximate the Hessian matrix of the log-likelihood function with respect to $\underline{\theta}$ yields expression (2.13) for the subvector of parameters $\underline{\beta}$, where the unknown elements of Ω_v are replaced by consistent estimates.

where $\underline{y}' = (\underline{y}_1', \underline{y}_2', \ldots, \underline{y}_T')$ and $\underline{y}_t = \underline{z}_t + \sum_{h=1}^{q} G_h \underline{v}_{t-h}$, after evaluation of the regressand, the regressors and the disturbance covariance matrix at consistent estimates of $\underline{\beta}$ and Ω_v, yields

$$(2.15.)\quad \hat{\hat{\underline{\beta}}} = [\hat{W}'\ \hat{M}'^{-1}\ (I_T \otimes \hat{\Omega}_v^{-1})\ \hat{M}^{-1}\hat{W}]^{-1}\ [\hat{W}'\ \hat{M}'^{-1}(I_T \otimes \hat{\Omega}_v^{-1})\ \hat{M}^{-1}\hat{\underline{y}}]$$

where "^" denotes that the quantities are evaluated at consistent estimates, for example $\hat{W} = (W_0, W_1, \hat{W}_2)$. The two-step estimator in (2.15.), which is similar to those proposed by Reinsel (1976) for other models, has the same asymptotic properties as the ML estimator ignoring possible restrictions on the parameters coming from the specification in (2.1). The estimator $\hat{\hat{\underline{\beta}}}$ is consistent, asymptotically normally distributed and efficient, with a large sample covariance matrix consistently estimated by

$$(2.16.)\quad \hat{V}(\hat{\hat{\underline{\beta}}}) = [\hat{W}'\hat{M}'^{-1}\ (I_T \otimes \hat{\Omega}_v^{-1})\ \hat{M}^{-1}\ \hat{W}]^{-1}.$$

Expression (2.15.) is in fact an approximation to the second step of the Newton-Raphson procedure. The approximation is such that it implements the second step of the Gauss-Newton* algorithm as is easily seen from writing

$$S = -\tfrac{1}{2}\ \underline{v}'\ (I_T \otimes \Omega_v^{-1})\ \underline{v} = \varepsilon'\varepsilon$$

where $\underline{\varepsilon} = P\underline{v}$ and P is the matrix obtained from the decomposition of the positive definite matrix $(I_T \times \Omega_v^{-1}) = P'P$.
The derivative of $\underline{\varepsilon}$ with respect to $\underline{\beta}$ is

$$(2.17.)\quad \frac{\partial \underline{\varepsilon}}{\partial \underline{\beta}} = -\ W'M'^{-1}P'$$

The second step of the Gauss-Newton procedure can be written as

$$(2.18.)\quad \hat{\hat{\underline{\beta}}} = \hat{\underline{\beta}} - [\frac{\partial \underline{\varepsilon}}{\partial \underline{\beta}}\ (\hat{\underline{\beta}}) \frac{\partial \underline{\varepsilon}}{\partial \underline{\beta}}\ (\hat{\underline{\beta}})']^{-1}\ \frac{\partial \underline{\varepsilon}}{\partial \underline{\beta}}\ (\hat{\underline{\beta}})\ \underline{\varepsilon}\ (\hat{\underline{\beta}})$$

$$= \hat{\underline{\beta}} - [\hat{W}'\hat{M}'^{-1}\ (I_T \otimes \hat{\Omega}_v^{-1})\ \hat{M}^{-1}\hat{W}]^{-1}\ \hat{W}'\hat{M}'^{-1}\ (I_T \otimes \hat{\Omega}_v^{-1})\underline{v}$$

which is - using expression (2.14.) - equivalent to (2.15.).

* The reader, who is not familiar with numerical procedures to solve systems of nonlinear equations, is referred to Goldfeld & Quandt (1972).

It should be obvious that the two-step Gauss-Newton approximation to the two-step Newton-Raphson* procedure is one member of a class of estimators which are asymptotically equivalent to the ML estimator. Further, expression (2.18.) can be calculated either using the analytical derivatives given in (2.17), evaluating them at consistent parameter estimates, as has been done e.g. by Nelson (1976), or by numerical calculation of the partial derivatives in (2.18.) of $\underline{\varepsilon}$ with respect to $\underline{\beta}$, as is proposed by Box and Jenkins (1970) for univariate models. Other estimators of $\underline{\beta}$ asymptotically equivalent to the ML estimator are characterized by a particular choice for the matrix $\Gamma(\hat{\underline{\beta}})$ in (2.13).

Joint estimation of the parameters in the system in (2.4.) involves the following steps:

1. Using the error representation shown in (2.7.), estimate the parameters of each equation separately using, for example, the Box-Jenkins (1970) non-linear least squares approach. The estimates so obtained will be consistent but not efficient. However the main objective of this first step is to obtain an estimate of $\underline{v}_t$, say $\hat{\underline{v}}_t$, t=1,2,...,T.

2. Use the $\hat{\underline{v}}_t$'s to form an estimate of the covariance matrix $\Omega_v = E\underline{v}_t\underline{v}_t'$, namely

$$(2.19.) \qquad \hat{\Omega}_v = \sum_{t=1}^{T} \hat{\underline{v}}_t\hat{\underline{v}}_t'/T$$

3. Express the i'th equation of the system (2.4.) as:

$$(2.20a.) \qquad z_{it} = c_i + \underline{w}_{1it}'\underline{\theta}_i + \underline{q}_t'\underline{\gamma}_i + v_{it} \qquad \begin{matrix} i=1,2,\ldots,p \\ t=1,2,\ldots,T \end{matrix}$$

where

$$\underline{w}_{1it}' = (z_{it-1}, z_{it-2}, \ldots, z_{it-n_i})$$

$$\underline{\theta}_i' = (\theta_{i1}, \theta_{i2}, \ldots, \theta_{in_i})$$

$$\underline{q}_t' = (\underline{v}_{t-1}', \underline{v}_{t-2}', \ldots, \underline{v}_{t-m})$$

$$\underline{\gamma}_i' = (\underline{\gamma}_{i1}', \underline{\gamma}_{i2}', \ldots, \underline{\gamma}_{im}')$$

with γ_{ij}' being the i'th row of G_j, j=1,2,...,m.
Expressing (2.20a.) for all t, we have

* The second order derivative of S with respect to β_i and β_j is

$$\frac{\partial^2 S}{\partial\beta_i\partial\beta_j} = 2\left[\frac{\partial^2\underline{\varepsilon}}{\partial\beta_i\partial\beta_j}\right]'\underline{\varepsilon} + 2\left[\frac{\partial\varepsilon}{\partial\beta_i}\right]'\left[\frac{\partial\varepsilon}{\partial\beta_j}\right] .$$

Under the assumptions underlying our model, it can be shown by using the strong law for martingales (see e.g. Feller (1966), p. 238) that the first r.h.s. term converges to zero in probability, so that the two-step estimator in (2.18) implements expression (2.13) and therefore has the same asymptotic properties as the ML estimator.

(2.20b.) $\underline{z}_i = W_{1i}\underline{\theta}_i + Q\underline{\gamma}_i + \underline{v}_i$ $\qquad i=1,2,\ldots,p$

where $\underline{w}'_{1it}$ and q'_t are typical rows of W_{1i} and Q respectively.

We then apply ordinary least squares to each equation (2.20a.), after replacing Q by $\hat{Q}$, a matrix of the first step residuals $\hat{\underline{v}}_t$, to obtain consistent estimates of $\underline{\gamma}_i$.

4. Compute expression (2.15.). Iteration of the steps 1-4 yields the ML estimator* given known and fixed initial conditions. In small samples it is not clear that the iterated estimator for $\underline{\beta}$ is to be preferred to $\hat{\hat{\underline{\beta}}}$. For example, it is well-known that ML estimators for parameters of many models have poor finite sample properties relative to usually employed loss functions**. Nelson (1976) provides Monto Carlo results pertaining to a system similar to a particular set of FEs that indicate a substantial gain of efficiency of the two-step estimators with respect to univariate procedures. The finite sample properties of $\hat{\hat{\underline{\beta}}}$ in (2.15) and estimators obtained by iteration are as yet not established.

The system of FEs (2.10.) can also be written as

(2.21.) $\underline{z} = W_0\underline{c} + W_1\underline{\theta} + W_2\underline{\gamma} + \underline{v}$

Generalized least squares applied to (2.21.) lead to

(2.22.) $\hat{\underline{\beta}}_{GLS} = [W' \, (I_T \otimes \Omega_v^{-1}) \, W']^{-1} \, W' \, (I_T \otimes \Omega_v^{-1})\underline{z}$

Since Ω_v and the elements of W_2, the lagged errors, usually have unknown values, (2.22.) cannot be computed. However, we can compute

(2.23.) $\hat{\hat{\underline{\beta}}}_{GLS} = [\hat{W}' \, (I_T \otimes \hat{\Omega}_v^{-1}) \, \hat{W}]^{-1} \, \hat{W}' \, (I_T \otimes \hat{\Omega}_v^{-1}) \, \underline{z}$

where $\hat{W}$ denotes that lagged error terms in W_2 have been replaced by their sample estimates. Expression (2.23.) gives a consistent joint estimator for $\underline{\beta}$, but it usually is not efficient. From a comparison with (2.15.), it is

* Other iterative algorithms that may be computationally more efficient can be employed to compute the ML estimate--- see e.g. Chow and Fair (1973) who considered a dynamic system with AR errors.

** See Zellner (1971) for some results relating to ML estimation of parameters of the log-normal distribution. For static simultaneous equation models, ML estimators frequently are found to possess no finite moments and hence have unbounded risk relative to a quadratic loss function. Last, Stein's well-known results indicate that ML estimators are often inadmissible relative to a quadratic loss function --- see references and analysis in Zellner and Vandaele (1974).

obvious that the estimator in (2.23.) is not a solution to the first order conditions for a maximum of the likelihood function, so that iterative solution of (2.23.) will not yield the ML estimator. (2.23.) is a solution of the first order conditions for a maximum of the likelihood function with respect to $\underline{\beta}$ under the condition that $W_2 = \hat{W}_2$ implying that $\partial S/\partial \underline{\beta}$ is linear in $\underline{\beta}$ (see e.g. Maddala (1971) for a similar discussion on single equation models).

As mentioned above, FEs are often encountered in which the $\underline{\theta}_i$ coefficient vectors in (2.8) are all the same, that is $\underline{\theta}_1 = \underline{\theta}_2 = \underline{\theta}_3 = \ldots = \underline{\theta}_p = \underline{\theta}^{(r)}$. In such cases, the restricted matrix W_1 in (2.9) takes the form

$$(2.24.) \qquad W_1^{(r)} = \begin{bmatrix} \underline{z}_0 & \underline{z}_{-1} & \cdot & \cdot & \underline{z}_{-p+1} \\ \underline{z}_1 & \underline{z}_{-2} & & & \underline{z}_{-p+2} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \underline{z}_{T-1} & \cdot & \cdot & \cdot & z_{T-p} \end{bmatrix}$$

We then write the system (2.10.) as:

$$(2.25.) \qquad \underline{z} = W_0 \underline{c} + W_1^{(r)} \underline{\theta}^{(r)} + \underline{u}$$

Then the approximate ML estimators for $\underline{c}$, $\underline{\gamma}$ and $\underline{\theta}^{(r)}$ are given by:

$$(2.26a.) \begin{pmatrix} \hat{\underline{c}} \\ \hat{\underline{\gamma}} \end{pmatrix} = [(W_0\hat{W}_2)' \, \hat{\Omega}^{-1} (W_0\hat{W}_2)]^{-1} \, (W_0\hat{W}_2) \, \hat{\Omega}^{-1}(\underline{\hat{y}} - W_1^{(r)}\hat{\underline{\theta}})$$

$$(2.26b.) \quad \hat{\underline{\theta}}^{(r)} = [W_1^{(r)} R W_1^{(r)}]^{-1} \, W_1^{(r)'} \, R\underline{\hat{y}}$$

with

$$R = \hat{\Omega}^{-1} - \hat{\Omega}^{-1}(W_0\hat{W}_2) \, [(W_0\hat{W}_2)' \, \hat{\Omega}^{-1}(W_0\hat{W}_2)]^{-1} \, (W_0\hat{W}_2)' \, \hat{\Omega}^{-1} \quad ,$$

where $\hat{\Omega}$ is an estimate of $\Omega = [M'(I_T \otimes \Omega_v)M]$ and $\hat{W}_2$ is formed using lagged residuals. The large sample covariance matrix for the restricted estimators in (2.26.), $(\hat{\underline{c}}, \hat{\underline{\theta}}^{(r)'}, \hat{\underline{\gamma}}')$, denoted V_r, is consistently estimated by

$$(2.27.) \qquad \hat{V}_r = \begin{bmatrix} W_0'\hat{\Omega}^{-1}W_0 & W_0'\hat{\Omega}^{-1}W_1^{(r)} & W_0'\hat{\Omega}^{-1}\hat{W}_2 \\ W_1^{(r)'}\hat{\Omega}^{-1}W_0 & W_1^{(r)'}\hat{\Omega}^{-1}W_1^{(r)} & W_1^{(r)'}\hat{\Omega}^{-1}\hat{W}_2 \\ \hat{W}_2'\hat{\Omega}^{-1}W_0 & \hat{W}_2'\hat{\Omega}^{-1}W_1^{(r)} & \hat{W}_2'\hat{\Omega}^{-1}\hat{W}_2 \end{bmatrix}$$

In the case of general linear restrictions on the elements of $\underline{\beta}$ in (2.14.), say $C\underline{\beta} = \underline{c}$, where C is a given matrix with q linearly independent rows of rank q, and $\underline{c}$ is a given qx1 vector, an estimator of $\underline{\beta}$ satisfying the restrictions is given by:

$$(2.28.) \quad \underline{\tilde{\hat{\beta}}} = \underline{\hat{\hat{\beta}}} - (\hat{W}'\hat{\Omega}^{-1}\hat{W})^{-1} C' \, [C(\hat{W}'\hat{\Omega}^{-1}\hat{W})^{-1} C']^{-1} \, (C\underline{\hat{\hat{\beta}}} - \underline{c})$$

with $\underline{\hat{\hat{\beta}}}$ as shown in (2.15.), $\hat{\Omega} = [\hat{M}'^{-1}(I_T \otimes \hat{\Omega}_v^{-1})\hat{M}^{-1}]$

and large sample covariance matrix, $V(\underline{\tilde{\beta}})$, consistently estimated by:

$$(2.29.) \quad \hat{V}(\underline{\tilde{\beta}}) = (\hat{W}'\hat{\Omega}^{-1}\hat{W})^{-1} - (\hat{W}'\hat{\Omega}^{-1}\hat{W})^{-1} C' \, [C(\hat{W}'\hat{\Omega}^{-1}\hat{W})^{-1}C']^{-1} \, C(\hat{W}'\hat{\Omega}^{-1}\hat{W})^{-1}.$$

While (2.28.) and (2.29.) are relevant for the case of general linear restrictions, it should be appreciated that the matrices involved in the expressions are quite large from a numerical point of view for systems even of moderate size.*

To test the restriction that $\underline{\theta}_1 = \underline{\theta}_2 = \ldots = \underline{\theta}_p = \underline{\theta}^{(r)}$, an nx1 vector, introduced in connection with (2.25.), we consider the following residual sums of squares:

$$(2.30.) \quad SS_r = (\underline{\hat{\hat{y}}} - \hat{\hat{W}}^{(r)} \, \underline{\hat{\hat{\beta}}}^{(r)})' \, \hat{\hat{\Omega}}^{-1} \, (\underline{\hat{\hat{y}}} - \hat{\hat{W}}^{(r)} \, \underline{\hat{\hat{\beta}}}^{(r)})$$

and

$$(2.31.) \quad SS_u = (\underline{\hat{\hat{y}}} - \hat{\hat{W}}\underline{\hat{\hat{\beta}}})' \, \hat{\hat{\Omega}}^{-1}(\underline{\hat{\hat{y}}} - \hat{\hat{W}}\underline{\hat{\hat{\beta}}})$$

where $\underline{\hat{\beta}}^{(r)'} = (\underline{\hat{c}}', \underline{\hat{\theta}}^{(r)'}, \underline{\hat{y}}')$ in (2.26.), $\hat{\hat{W}}^{(r)} = (W_0 \;\; W_1^{(r)} \; \hat{\hat{W}}_2)$,

$\hat{\hat{\Omega}} = \hat{\hat{M}}'^{-1}(I_T \otimes \hat{\hat{\Omega}}_v)\hat{\hat{M}}^{-1}$ and $\underline{\hat{\hat{\beta}}}$ in (2.31) is given in (2.15), with "$\hat{\hat{}}$" denoting that the quantities are computed using the second step estimates of $\underline{\beta}$ and $\underline{\beta}^{(r)}$. Thus, the approximate likelihood ratio

$$(2.32.) \quad N \log (SS_r/SS_u)$$

is in large samples χ^2_m distributed where N = Tp and m= n(p-1), the number of restrictions involved in $\underline{\theta}_1 = \underline{\theta}_2 = \ldots = \underline{\theta}_p = \underline{\theta}^{(r)}$. Thus (2.32.) provides a large sample χ^2 test of a frequently encountered hypothesis in model construction. In a similar fashion, large sample χ^2 tests of the general linear hypothesis $C\underline{\beta} = \underline{c}$ can be constructed.

* With respect to the large matrices that are encountered in joint estimation procedures and that will usually lead to a multicollinearity problem, it is worthwhile to mention the use of approximate Bayes estimates such as considered by Zellner and Vandaele (1974). In a Monte Carlo study of the small sample (T=20) properties of several estimators for a dynamic model with first order autoregressive errors, Swamy and Rappoport (1978) conclude that in terms of mean square errors the ridge regression and the approximate minimum mean square error estimates of the regression coefficients are significantly better than alternative estimates such as ML or Hatanaka's (1974) residual adjusted estimates.

III. Specification of and Estimation and Testing Procedures for Sets of Transfer Functions

To specify a set of transfer functions (TFs), we partition the vector $\underline{z}_t$ in (2.1.) as follows, $\underline{z}_t' = (\underline{y}_t' \ \underline{x}_t')$ where $\underline{y}_t$ is a p_1x1 vector of endogenous variables and $\underline{x}_t$ is a p_2x1 vector of exogenous variables with $p_1 + p_2 = p$. With $\underline{z}_t$ so partitioned, the system in (2.1.) becomes:

$$(3.1.) \quad \begin{pmatrix} H_{11}(L) & H_{12}(L) \\ H_{21}(L) & H_{22}(L) \end{pmatrix} \begin{pmatrix} \underline{y}_t \\ \underline{x}_t \end{pmatrix} = \begin{pmatrix} \bar{\underline{c}}_1 \\ \bar{\underline{c}}_2 \end{pmatrix} + \begin{pmatrix} F_{11}(L) & F_{12}(L) \\ F_{21}(L) & F_{22}(L) \end{pmatrix} \begin{pmatrix} \underline{e}_{1t} \\ \underline{e}_{2t} \end{pmatrix}$$

The assumption that $\underline{x}_t$ is exogenous gives rise to the following restrictions on the system in (3.1.):

$$(3.2.) \quad H_{21}(L) \equiv 0, \quad F_{12}(L) \equiv 0, \quad \text{and } F_{21}(L) \equiv 0.$$

With the restrictions in (3.2.) imposed on (3.1.), we have

$$(3.3a.) \quad H_{11}(L)\underline{y}_t + H_{12}(L)\underline{x}_t = \bar{\underline{c}}_1 + F_{11}(L)\underline{e}_{1t}$$

$$(3.3b.) \quad H_{22}(L)\underline{x}_t = \bar{\underline{c}}_2 + F_{22}(L)\underline{e}_{2t}.$$

The system in (3.3a.) is in the form of a set of linear, dynamic simultaneous equations while that in (3.3b.) is a set of ARMA equations for the exogenous variables.

The TFs associated with (3.3a.), obtained by multiplying both sides of (3.3a.) by $H_{11}^*(L)$, the adjoint matrix associated with $H_{11}(L)$ are

$$\begin{aligned} (3.4.) \quad |H_{11}(L)|\underline{y}_t &= \underline{c}_1 - H_{11}^*(L)H_{12}(L)\underline{x}_t + H_{11}^*(L)F_{11}(L)\underline{e}_{1t} \\ &= \underline{c}_1 + \Delta(L)\underline{x}_t + K(L)\underline{e}_{1t} \end{aligned}$$

where $|H_{11}(L)|$ is the determinant of $H_{11}(L)$, $\underline{c}_1$ is a p_1x1 vector of constants, $\Delta(L) \equiv - H_{11}^*(L)H_{12}(L) = \sum_{i=0}^{r} \Delta_i L^i$ and $K(L) \equiv H_{11}^*(L)F_{11}(L) = \sum_{j=0}^{q} K_j L^j$. In order to identify (3.4.), we assume that $|H_{11}(L)|$ and $|K(L)|$ have their roots outside the unit circle and that the r.h.s. and the l.h.s. of (3.4.) have no factors in common. The ith equation of (3.4.) is given by:

(3.5a.) $\phi(L)y_{it} = c_{1i} + \underline{\delta}_i' \; \underline{x}_t + \underline{k}_i'\underline{e}_{1t}$

where $\phi(L) \equiv |H_{11}(L)|$, $\underline{\delta}_i'$ is the i'th row of $\Delta(L)$ and $\underline{k}_i'$ is the i'th row of $K(L)$. Just as with the FEs, the AR polynomial $\phi(L)$ is the same in each equation if no cancelling occurs. To allow for possibly different $\phi(L)$ in different equations, we shall write the TF system* as:

(3.5b.) $\phi_i(L)y_{it} = c_{1i} + \underline{\delta}_i'\underline{x}_t + \underline{k}_i'\underline{e}_{1t}, \qquad i=1,2,\ldots,p_1,$

with $\phi_i(L) = 1 - \phi_{i1}L - \phi_{i2}L^2 - \ldots \phi_{im_i}L^{m_i}$, with m_i assumed known. Since the error terms in (3.5b.) have a structure similar to those in (2.4.), the representations presented in (2.5.) and (2.7.) are relevant here.

Each TF can be estimated separately using single equation non-linear least squares or the single equation ML procedure. Joint estimation of the parameters of the set of TFs (3.5.) will now be considered. Therefore we write the system (3.5.) as

(3.6.) $\underline{y}_t = Y_t\underline{\phi} + X_{0t}\underline{c}_1 + X_t\underline{\delta} + \underline{u}_{1t}$

where

$$\underset{p_1 \times \sum_{i=1}^{p_1} m_i}{Y_t} = \begin{bmatrix} y_{1t-1} & y_{1t-2} & \cdots & y_{1t-m_1} & 0 & \cdots & & 0 \\ 0 & & & 0 & y_{2t-1} & \cdot & y_{2t-m_2} & \cdots\; 0 \\ \vdots & & & & & & & \\ \vdots & & & & & & & 0 \\ 0 & \cdots & & & & & & y_{p_1t-m_{p_1}} \end{bmatrix}$$

$\underline{\phi}' = [\phi_{11} \;\; \phi_{12} \;\; \ldots \;\; \phi_{1m_1} \;\; \phi_{21} \; \ldots \; \phi_{2m_2} \; \ldots \; \phi_{p_1m_{p_1}}]$

$X_{0t} = I_{p_1}$

* We implicitly assume that the TF system has been normalized.

$$X_t = \begin{bmatrix} x_{1t} & x_{1t-1} & \cdots & x_{1t-r_{11}} & 0 \cdots 0 & x_{2t} & x_{2t-2} & \cdots & x_{2t-r_{21}} & \cdots & x_{p_2t-r_{p_21}} & 0 \cdots 0 \\ x_{1t} & x_{1t-1} & \cdots & x_{1t-r_{12}} & 0 \cdots 0 & x_{2t} & \cdots & & & & x_{p_2t-r_{p_22}} & \\ \vdots & & & & & & & & & & & \\ x_{1t} & & & & & & & & & & x_{p_2t-r_{p_2p_1}} & \end{bmatrix}$$

$$p x (p_2 + \sum_{i=1}^{p_2} \max_{j=1,\ldots,p_2} r_{ij})$$

$\underline{\delta}' = \text{vec } [\Delta_1 \; \Delta_2 \ldots \Delta_r]'$ and, provided the values of the initial conditions are zero, $\underline{u}_1 = K \, \underline{\varepsilon}_1$, where $\underline{\varepsilon}_1' = [(K_0\underline{e}_{11})' \;\; (K_0\underline{e}_{12})' \ldots (K_0\underline{e}_{1T})']$, $E(\underline{\varepsilon}_{1t} \, \underline{\varepsilon}_{1t}') = K_0 \, K_0' = \Omega_\varepsilon$, and

$$K = \begin{bmatrix} I_{p_1} & 0 & \cdots & & 0 \\ & & & & \vdots \\ K_1K_0^{-1} & \ddots & & & 0 \\ \vdots & & \ddots & & \vdots \\ K_qK_0^{-1} & & & \ddots & \vdots \\ 0 & & & & \vdots \\ 0 \cdots & K_qK_0^{-1} & K_1K_0^{-1} & & I_{p_1} \end{bmatrix}$$

For a sample of T observations, we write the system (3.6.) as

$$(3.7.)\qquad \underline{y} = Y\underline{\phi} + X_0\underline{c}_1 + X\underline{\delta} + \underline{u}_1$$

$$= Z_1\underline{\lambda}_1 + \underline{u}_1$$

with $\quad Z_1 = (Y, X_0, X)$

$\qquad\quad \underline{\lambda}_1' = (\underline{\phi}', \underline{c}_1', \underline{\delta}')$

As in the preceding section, the likelihood function for the unrestricted system of TFs, conditional on zero starting values*, can be written as:

$$(3.8.)\qquad L(\underline{\lambda}_1, \Omega_\varepsilon, K, \underline{y}) \propto |\Omega_\varepsilon|^{-T/2} \exp(S)$$

where $\quad S = -\frac{1}{2}\,(\underline{y}-Z_1\underline{\lambda}_1)'K'^{-1}\,(I_T \otimes \Omega_\varepsilon^{-1})\,K^{-1}(\underline{y}-Z_1\underline{\lambda}_1)$

The first order conditions for a maximum of the log-likelihood functions are

$$(3.9.)\qquad \frac{\partial S}{\partial \underline{\lambda}} = Z'K'^{-1}(I_T \otimes \Omega_\varepsilon^{-1})\,K^{-1}\underline{u}_1 = \underline{0}$$

where $\quad \underline{\lambda}' = (\underline{\lambda}_1', \underline{\lambda}_2')$, with $\underline{\lambda}_2' = \text{vec}\,[K_1K_0^{-1}, \ldots, K_qK_0^{-1}]'$

and $\quad Z = [Z_1, Z_2]$

$Z_2' = [\underline{\varepsilon}_{ij}^{\ell'}]$, where $i=1,2,..,p_1$, $\ell=1,2,..,q$, $j=1,2,..,p_1$, and with typical row

$$\underline{\varepsilon}_{ij}^{\ell'} = [\underbrace{0,0,\ldots\,0}_{\ell p_1 \text{ times}}\;,\;\underbrace{0,\ldots\,0\;\varepsilon_{j1}\;0..0}_{p_1 \text{ elements}}\;\ldots\ldots\;\underbrace{0\ldots\,0\;\varepsilon_{jT-\ell}\;0\ldots 0}_{p_1 \text{ elements}}]$$

i'th position

As in to the preceding section, a fully** efficient two-step estimator of $\underline{\lambda}$ is obtained using expression (2.13.) to yield

$$(3.10.)\qquad \hat{\hat{\underline{\lambda}}} = [\hat{Z}'\,\hat{K}'^{-1}\,(I_T \otimes \hat{\Omega}_\varepsilon^{-1})\,\hat{K}^{-1}\hat{Z}]^{-1}\,[\hat{Z}'\hat{K}'^{-1}\,(I_T \otimes \hat{\Omega}_\varepsilon^{-1})\,\hat{K}^{-1}\,\hat{\underline{w}}]$$

* As with the set of FEs, the starting values may be "backforecast" using single TF equations. They may also be considered as unknown parameters to be estimated (e.g. Phillips (1966)).

** This means, that the estimator is as efficient as the ML estimator for the parameters of the TF form, ignoring restrictions coming from the underlying structural form.

where "^" denotes that the unobserved quantities are computed at consistent estimates of $\underline{\lambda}$, $\hat{\underline{\lambda}}$, and the sample residuals obtained from the estimates and

$$\underline{w}' = (\underline{w}_1', \ldots, \underline{w}_T'), \text{ where } \underline{w}_t = \underline{y}_t + \sum_{h=1}^{q} K_h K_0^{-1} \underline{\varepsilon}_{1t-h} .$$

The remarks made in the preceding section concerning expression (2.15.) also apply to (3.10.). The steps in obtaining joint estimates of the parameters in (3.5b.) are as follows:

1. Fit individual equations of (3.5b.) to obtain consistent estimates of the contemporaneous residuals, $\hat{\underline{\varepsilon}}_t$, where $\underline{\varepsilon}_t = K_0 \underline{e}_{1t}$.
2. Use the residuals to form a consistent estimate of the contemporaneous covariance matrix of $\underline{\varepsilon}_t$, $E\underline{\varepsilon}_t\underline{\varepsilon}_t' = K_0K_0' \equiv \Omega_\varepsilon$, namely

$$\text{(3.11.)} \quad \hat{\Omega}_\varepsilon = \sum_{t=1}^{T} \hat{\underline{\varepsilon}}_t \hat{\underline{\varepsilon}}_t' / T.$$

3. Rewrite the i'th equation of (3.5b.) as:

$$\text{(3.12a.)} \quad y_{it} = c_{1i} + \sum_{\ell=1}^{m_i} y_{i,t-\ell}\phi_{i\ell} + \sum_{\ell=0}^{r_{i1}} x_{1t-\ell}\delta_{i1\ell} + \ldots + \sum_{\ell=0}^{r_{ip_2}} x_{p_2,t-\ell}\delta_{ip_2\ell}$$

$$+ \underline{\varepsilon}_{t-1}'\underline{\psi}_{i1} + \underline{\varepsilon}'_{t-2}\underline{\psi}_{i2} + \ldots + \underline{\varepsilon}_{t-q_i}'\underline{\psi}_{i,q_i} + \varepsilon_{it}$$

or

$$\text{(3.12b.)} \quad \underline{y}_i = c_{1i}\underline{\iota} + Y_i\underline{\phi}_i + X_i\underline{\gamma}_i + \xi_i\underline{\psi}_i + \underline{\varepsilon}_i$$

$$= J_i\underline{\eta}_i + \underline{\varepsilon}_i \qquad i=1,2,\ldots,p_1$$

where $\underline{y}_1$ has typical element y_{it},
Y_1 has typical row $(y_{it-1}, y_{it-2}, \ldots, y_{it-m_i})$,
X_i has typical row $(x_{1t}, x_{1t-1}, \ldots, x_{1t-r_{1i}}, \ldots, x_{p_2t}, x_{p_2t-1}, \ldots, x_{p_2t-r_{p_2i}})$,
ξ_i has typical row $(\underline{\varepsilon}_{t-1}', \underline{\varepsilon}'_{t-2}, \ldots, \underline{\varepsilon}_{t-q_i}')$,
$\underline{\varepsilon}_i$ has typical element ε_{it},
$\underline{\phi}_i' = (\phi_{i1}, \phi_{i2}, \ldots, \phi_{im_i})$,
$\underline{\delta}_i' = (\delta_{i10}, \delta_{i11}, \ldots, \delta_{i1r_{1i}}, \ldots, \delta_{ip_20}, \delta_{ip_21}, \ldots, \delta_{ip_2r_{p_2i}})$,
$\underline{\psi}_i' = (\underline{\psi}_{i1}', \underline{\psi}_{i2}', \ldots, \underline{\psi}_{i,q_i}')$,
$J_i = (\underline{\iota} \vdots Y_i \vdots X_i \vdots \underline{\varepsilon}_i)$
$\underline{\iota}' = (1,1,\ldots,1)$ Tx1
$\underline{\eta}_i' = (c_{1i} \vdots \underline{\phi}_i' \vdots \underline{\gamma}_i' \vdots \underline{\psi}_i')$

We then apply ordinary least squares to each equation (3.12b.), after replacing $\underline{\varepsilon}_i$ by $\hat{\underline{\varepsilon}}_i$, a matrix of the first step residuals $\hat{\underline{\varepsilon}}_t$ to obtain consistent estimates of $\underline{\psi}_i$.

4. Compute expression (3.10.), using the residuals obtained in step 2 and the consistent parameter estimates $\hat{\Omega}_\varepsilon$ and $\hat{\underline{\eta}}_i$ to evaluate the unknown quantities in (3.10.). Iteration of (3.10.) yields the ML estimator given known and fixed initial conditions. To compute the inverse of $\hat{K}$, one ought to exploit the block-triangular structure of this matrix. This reduces the inversion of a Tp_1 matrix to addition and multiplication of $p_1 x p_1$ matrices, as discussed in Palm (1977.b.).

The large sample covariance matrix of the estimator proposed in (3.10.) is consistently estimated by

$$\text{(3.13.)} \quad \hat{\text{var}}\,(\hat{\underline{\lambda}}) = [\hat{Z}'\hat{K}'^{-1}\,(I_T \otimes \hat{\Omega}_\varepsilon^{-1})\,\hat{K}^{-1}\hat{Z}]^{-1} \quad .$$

In considering TF estimation, it is important to realize that the number of parameters in each TF can be large when there are several input variables in the vector $\underline{x}_t$ and lags relating to them are long. In such cases it will be expedient to consider reducing the number of free parameters to be estimated by making assumptions regarding the forms of lagged responses as is done in the distributed lag literature --- see e.g. Almon (1965), Dhrymes (1971), Shiller (1973) and Zellner (Ch. 7, 1971). And of course, introducing the restriction that $\phi_1(L) \equiv \phi_2(L) \equiv \ldots \equiv \phi_{p_1}(L) \equiv \phi(L)$ in (3.5b.), when warranted, will lead to fewer free parameters to be estimated. A large sample χ^2 test of the hypothesis that $\phi_1(L) \equiv \phi_2(L) \equiv \ldots \equiv \phi_{p_1}(L) \equiv \phi(L)$ can be constructed as in the case of FEs where a similar hypothesis was considered.

IV. Estimation and Testing Procedures for Structural Equations

The structural equations (SEs), shown explicitly in (3.3a.), will not be considered. We shall assume that a sufficient number of zero restrictions has been imposed on the parameters of the system such that the remaining free parameters are identified --- see Hannan (1971) and Hatanaka (1975). In what follows, we shall first take up "single-equation" estimation techniques for parameters in individual SEs and then go on to develop a "joint estimation" procedure that can be employed to estimate parameters appearing in a set of SEs.

4.1. Single-Equation Estimation Procedure

The i'th SE of the system in (3.3a.) is given by:

$$(4.1.) \qquad \sum_{j=1}^{p_1} h_{ij}(L)y_{jt} + \sum_{j=p_1+1}^{p} h_{ij}(L)x_{jt} = \bar{c}_{1i} + \sum_{j=1}^{p_1} f_{ij}(L)e_{jt}$$

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

On imposing identifying zero restrictions and a normalization rule, $h_{ii0}=1$*, the remaining free parameters of (4.1.) can be estimated utilizing the techniques described below.

As shown in connection with (2.7.) above, we can write

$$(4.2.) \qquad \sum_{j=1}^{p_1} f_{ij}(L)e_{jt} = \phi_i(L)\xi_{it}$$

where $\phi_i(L)\xi_{it} = \xi_{it} + \phi_{i1}\xi_{it-1} + \phi_{i2}\xi_{it-2} +..+ \phi_{iq}\xi_{it-q_i}$,

where $q_i = \max_j q_{ij}$, with q_{ij} the degree of $f_{ij}(L)$, and ξ_{it} is a non-autocorrelated, normally distributed disturbance term with zero mean and constant finite variance, σ_ξ^2 , for all t. On substituting form (4.2.) in (4.1.), we have

$$(4.3.) \qquad \sum_{j=1}^{p_1} h_{ij}(L)y_{jt} + \sum_{j=p_1+1}^{p} h_{ij}(L)x_{jt} = \bar{c}_{1i} + \phi_i(L)\xi_{it} \quad , t=1,2,\ldots,T$$

with the property that x_{jt} and $\xi_{it'}$ are independent for all j, t and t'.

Since more than one current endogenous variable appears in (4.3.), along with lagged endogenous variables, and since the disturbance terms are serially correlated, it is well-known that usual estimation techniques such as two-stage least-squares, etc., yield inconsistent structural coefficient estimates.

* Note $h_{ij}(L) \equiv h_{ij0} + h_{ij1}L + h_{ij2}L^2 + \ldots$.

Similarly, non-linear techniques for minimizing $\sum_{t=1}^{T}\xi_{it}^2$ with respect to the parameters of (4.3.) yield inconsistent coefficient estimates because of "simultaneous equation" complications*.

To get consistent estimates of the parameters in (4.3.), one can use an instrumental variables method using as instruments for the current and lagged endogenous variables, the current and lagged exogenous variables. One can also use the $y_{jt-q_i-\ell}$, $j=1,\ldots,p_1$, $\ell=1,2,\ldots$ as instruments for the current and lagged endogenous variables, as these instruments are independent of the error term $\phi_i(L)\xi_{it}$. The use of lagged endogenous variables as instruments has been proposed by Phillips (1966). On de basis of the instrumental variables estimates, $\hat{\hat{c}}_{1i}$ and $\hat{h}_{ij}$ (L) $j=1,2,\ldots p$ in (4.3.), one can compute the residuals.

$\hat{\eta}_{it}=\sum_{j=1}^{p_1}\hat{h}_{ij}(L)y_{jt}+\sum_{j=p_1+1}^{p}\hat{h}_{ij}(L)\,x_{jt}+\hat{\hat{c}}_{1i}$ and then fit a q_i' th order MA model to the residuals to get consistent estimates of the $\phi_{i\ell}$'s $\ell=1,2,\ldots q_i$.

Alternatively, as explained in Zellner and Palm (1974), one may use the FEs (or TFs) to substitute for current endogenous variables appearing in (4.3.) with coefficients with unknown values, that is y_{jt}, $j=1,2,\ldots,p_1$, for $j\neq i$. For example, the FEs for the y_{jt} given in (2.4.) are:

$$(4.4.)\qquad y_{jt}=c_j+\bar{\theta}_j(L)y_{jt}+\underline{\alpha}_j'\underline{e}_t$$

where $\bar{\theta}_j(L)$ is the homogeneous part of $-\theta_j(L)$. On substituting from (4.4.) in (4.3.) for y_{jt}, $j=1,2,\ldots,p_1$, for $j\neq i$, we obtain:

$$(4.5.)\qquad y_{it}+\sum_{\substack{j=1\\ j\neq i}}^{p_1}h_{ij0}\,[c_j+\bar{\theta}_j(L)y_{jt}+\underline{\alpha}_j'\underline{e}_t]+\sum_{j=1}^{p_1}\bar{h}_{ij}(L)y_{jt}+$$

$$+\sum_{j=p_1+1}^{p}h_{ij}(L)x_{jt}=\tilde{c}_{1i}+\theta_i(L)\xi_{it},\qquad t=1,2,\ldots T,$$

where $\bar{h}_{ij}(L)$ is the homogeneous part of $h_{ij}(L)$. On rearranging terms in (4.5.), we can express (4.5.) as**

* Explicitly the pdf for $\underline{\xi}_i'=(\xi_{i1},\xi_{i2},\ldots,\xi_{it},\ldots,\xi_{iT})$ is: $p(\underline{\xi}|\sigma_\xi^2)=(2\pi\sigma_\xi^2)^{-T/2}\exp\{-\underline{\xi}_i'\underline{\xi}_i/2\sigma_\xi^2\}$. This last expression, however, is not the likelihood function and thus minimization of $\sum_{t=1}^{T}\xi_{it}^2$ does not provide consistent maximum likelihood estimates. The difficulty is analogous to that arising in application of OLS to estimate structural parameters in usual structural equations.

** One has to be cautious that the regressor matrix does not become singular, as one substitutes linear combinations of lagged endogenous variables for the current endogenous variables.

$$(4.6.)\qquad y_{it} + \sum_{\substack{j=1\\ j\neq i}}^{p_1} h_{ij0}\,[c_j + \bar{\theta}_j(L)y_{jt}] + \sum_{j=1}^{p_1} \bar{h}_{ij}(L)y_{jt}$$

$$+ \sum_{j=p_1+1}^{p} h_{ij}(L)x_{jt} = \bar{c}_{1i} + \psi_i(L)v_{it}$$

where we have introduced $\psi_i(L)v_{it} = \phi_i(L)\xi_{it} - \sum_{\substack{j=1\\ j\neq i}}^{p_1} h_{ij0}\underline{\alpha}_j'\underline{e}_t$.

$\psi_i(L)$ is a polynomial lag operator of degree $r = \{\max q_i, r_j \mid j = 1,2,\dots,p_1; j\neq i\}$ with r_j = degree of the highest degree polynomial in the vector $\underline{\alpha}_j'$. The error terms, v_{it}, $t=1,2,\dots,T$ are normally and independently distributed each with zero mean and finite, common variance, $\sigma^2_{v_i}$.

Note that for given values of c_j and the parameters in $\bar{\theta}_j(L)$, (4.6.) is in the form of a TF that is linear in the parameters. In view of this our estimation approach involves analyzing (4.6.) as a TF with $c_j = \hat{c}_j$ and $\bar{\theta}_j(L) = \hat{\bar{\theta}}_j(L)$, where $\hat{c}_j$ and $\hat{\bar{\theta}}_j(L)$ are consistent estimates obtained from estimation of the FEs in (4.4.) Since the Jacobian of the transformation form the v_{it}'s to the y_{it}'s in (4.6.) is equal to 1, the likelihood function is given by:

$$(4.7.)\qquad p(\underline{y}_i|\underline{\zeta},X,\underline{y}_0) = (2\pi\sigma^2_{v_i})^{-T/2}\exp -\{\sum_{t=1}^{T} v^2_{it}/2\sigma^2_{v_i}\}$$

where $\underline{y}_i' = (y_{i1}, y_{i2}, \dots, y_{it}, \dots, y_{iT})$, $\underline{\zeta}$ denotes the vector of free parameters to be estimated in (4.6.), X is the matrix of observation on all exogenous variables, including initial values, $\underline{y}_0$ denotes initial values of the endogenous variables appearing in (4.6.), and

$$(4.8.)\qquad v_{it} = y_{it} - \bar{c}_{1i} + \sum_{\substack{j=1\\ j\neq i}}^{p_1} h_{ij0}\,[c_j + \bar{\theta}_j(L)y_{jt}] + \sum_{j=1}^{p_1} h_{ij}(L)y_{jt}$$

$$+ \sum_{j=p_1+1}^{p} h_{ij}(L)x_{jt} + \sum_{j=1}^{r} \psi_{ij}v_{it-j}.$$

Given that consistent estimates of c_j and $\bar{\theta}_j(L)$ are inserted in (4.8.), a non-linear computational algorithm, e.g. Marquardt's, can be utilized to obtain consistent estimates of the remaining free parameters of (4.6.). The inverse of the information matrix, evaluated at the consistent estimates, provides large sample standard errors. These results in conjunction with the large sample normal distribution of the estimates provide a basis for performing large sample tests of hypotheses.

The above procedure for estimating parameters of (4.6.) can be applied for $i=1,2,\ldots,p_1$ to obtain "single-equation" parameter estimates and residuals, $\hat{v}_{it}$, $i=1,2,\ldots,p_1$, and $t=1,2,\ldots,T$. Since

$v_{it} = \xi_{it} - \sum_{\substack{j=1\\j\neq i}}^{p_1} h_{ij0}\underline{\alpha}'_{j0}\underline{e}_t$, where $\underline{\alpha}'_{j0}$ is a vector of constants,

$$(4.9.)\qquad \hat{\xi}_{it} = \hat{v}_{it} + \sum_{\substack{j=1\\j\neq i}}^{p_1} \hat{h}_{ij0}\widehat{\underline{\alpha}'_{j0}\underline{e}_t} \qquad\qquad i=1,2,\ldots,p_1$$

is a consistent estimate of ξ_{it} where $\widehat{\underline{\alpha}'_{j0}\underline{e}_t}$ denotes a vector of residuals from the FE system in (4.4.) Also, from (3.3a.) and (4.3.), $\underline{\xi}_t = F_{110}\underline{e}_{1t} = \underline{u}_t$, where $\underline{\xi}'_t = (\xi_{1t},\xi_{2t},\ldots,\xi_{p_1t})$, $\underline{u}'_t = (u_{1t},u_{2t},\ldots,u_{p_1t})$ and

$$(4.10.)\qquad F_{11}(L)\underline{e}_{1t} = F_{110}\underline{e}_{1t} + F_{111}\underline{e}_{1t-1} + \ldots + F_{11r}\underline{e}_{1t-r} =$$

$$= \underline{u}_t + R_1\underline{u}_{t-1} + \ldots + R_r\underline{u}_{t-r}$$

with $R_1 = F_{111}F_{110}^{-1}$,.... , $R_r = F_{11r}F_{110}^{-1}$. The error vector $\underline{u}_t$ in (4.10.) is normally distributed with zero mean and covariance matrix $E\underline{u}_t\underline{u}'_t$ = $F_{110}F'_{110} = \Sigma$, a $p_1 x p_1$ pds matrix, for all t. Also $E\underline{u}_t\underline{u}'_{t'}=0$ for $t\neq t'$ and all t, t'. Thus from (4.9.), it is possible to compute $\hat{\underline{u}}_t = \hat{\underline{\xi}}_t$ once all equations of the system are estimated. The $\hat{\underline{u}}_t$'s thus computed will play a role in the joint estimationprocedure to be described in the next section.

4.2. Joint Estimation of a Set of Structural Equations

We now consider (4.1.) for $i=1,2,\ldots,p_1$,

$$(4.11.)\qquad \sum_{\ell=0}^{s} H_{11\ell}\underline{y}_{t-\ell} + \sum_{j=0}^{r} H_{12j}\underline{x}_{t-j} = \sum_{h=0}^{q} F_{11h}\underline{e}_{1t-h}$$

where the diagonal elements of H_{110} are equal to one. For a sample of T observations, we can write the system (4.11.) as

$$(4.12.)\qquad \underline{y} = Z_1\underline{\eta}_1 + F\underline{v} = Z_1\underline{\eta}_1 + \underline{u}$$

where $\qquad \underline{y}' = (\underline{y}'_1,\underline{y}'_2,\ldots,\underline{y}'_T)$

$$\underset{p_1 T x k}{Z_1} = (X_0 \quad Y_1 \quad X_1)$$

$$\underset{p_1 \times p_1 T}{X_0'} = [I_{p_1} \; I_{p_1} \; \cdots \; I_{p_1}]$$

$$\underset{p_1 T \times p_1^2}{Y_1} = (Y \otimes I_{p_1})$$

$$\underset{T \times p_1}{Y} = \begin{bmatrix} y_{11} & y_{21} & \cdots & y_{p_1 1} \\ y_{12} & & & . \\ . & & & . \\ y_{1T} & \cdots & \cdots & y_{p_1 T} \end{bmatrix}$$

$$X_1 = \begin{bmatrix} I_{p_1} \otimes (\underline{y}_0' \; \underline{y}_{-1}' \; \cdots \; \underline{y}_{-s+1}' \; \underline{x}_1' \; \underline{x}_0' \; \cdots \; \underline{x}_{-r+1}') \\ I_{p_1} \otimes (\underline{y}_1' \; \underline{y}_0' \qquad\qquad\qquad\qquad \underline{x}_{-r+2}') \\ \\ I_{p_1} \otimes (\underline{y}_{T-1}' \; \cdots \; \underline{y}_{T-s}' \; \underline{x}_T' \; \cdots \; x_{T-r}') \end{bmatrix}$$

$$p_1 T \times p_1(s+r+1)$$

$k = [p_1 + p_1^2 + p_1(s+r+1)]$*

$\underline{n}_1' = (\underline{\bar{c}}_1', \underline{n}_2', \underline{n}_3')$

$\underline{n}_2' = \text{vec}\,(I - H_{110})$

$\underline{n}_3' = \text{vec}\,(-H_{111} - H_{112} \cdots - H_{11s} - H_{120} \cdots - H_{12r})'$

* Of course, one has to impose restriction on the structural parameters, e.g.zero restrictions, for which some of the columns of Z_1 will be deleted.

$$
\underset{p_1T\times p_1T}{F} = \begin{bmatrix}
I_{p_1} & 0 & \cdots & \cdots & \cdots & 0 \\
F_{111}F_{110}^{-1} & \ddots & & & & \vdots \\
\vdots & \ddots & \ddots & & & \vdots \\
F_{11q}F_{110}^{-1} & & \ddots & \ddots & & \vdots \\
0 & \ddots & & \ddots & \ddots & 0 \\
0 \;\cdots\; 0 & F_{11q}F_{110}^{-1} & \cdots & & F_{111}F_{110}^{-1} & I_{p_1}
\end{bmatrix}
$$

$\underline{v}' = (\underline{v}_1' \; \ldots \; \underline{v}_T')$, $\underline{v}_t = F_{110}\underline{e}_{1t}$

with $E(\underline{v}_t) = 0$, $E(\underline{v}_t\underline{v}_{t'}') = \delta_{tt'}$ $F_{110}F_{110}' = \delta_{tt'}\Sigma_v$, where $\delta_{tt'}$ is the Kronecker delta.

Since the $\underline{v}_t$'s are assumed to be normally distributed, the likelihood function can be written as $L(\underline{y}, Z_1, \underline{\eta}_1, \Sigma_v, F) \propto$

$$
(4.13.) \quad |H_{110}|^T \, |\Sigma_v|^{-T/2} \exp\{-\tfrac{1}{2}[(\underline{y} - Z_1\,\underline{\eta}_1)' \, F'^{-1} \, (I_T \otimes \Sigma_v^{-1}) \, F^{-1}(\underline{y} - Z_1\,\underline{\eta}_1)]\}
$$

In order to keep the block-triangular structure of the matrix F, we proceed in a way slightly different from Reinsel (1977) and write the first order conditions for a maximum of the likelihood function as:

$$
(4.14a.) \quad \frac{\partial \ln L}{\partial \, \underline{\bar{c}}_1} = Z_0' \, F'^{-1} \, (I_T \otimes \Sigma_v^{-1}) \, F^{-1}\underline{u} = \underline{0}
$$

$$
(b) \quad \frac{\partial \ln L}{\partial \, \underline{\eta}_2} = -T \operatorname{vec}(H_{110})^{-1} + Y'F'^{-1} \, (I_T \otimes \Sigma_v^{-1}) \, F^{-1}\underline{u} = \underline{0}
$$

$$
(c) \quad \frac{\partial \ln L}{\partial \, \underline{\eta}_3} = X_1' \, F'^{-1} \, (I_T \otimes \Sigma_v^{-1}) \, F^{-1}\underline{u} = \underline{0}
$$

$$
(d) \quad \frac{\partial \ln L}{\partial \, \underline{\beta}} = X_2' \, F'^{-1} \, (I_T \otimes \Sigma_v^{-1}) \, F^{-1}\underline{u} = \underline{0}
$$

where $\underline{\beta}' = \text{vec}\,[F_1 \;\; F_2 \;\; \ldots \;\; F_q]'$

$X_2' = [\underline{v}_{ij}^{\ell'}\,]$, $i=1,2,\ldots,p_1$, $\ell = 1,2,\ldots,q$, $j=1,2,\ldots,p_1$

with a typical row

$$\underline{v}_{ij}^{\ell'} = [\underbrace{0,0 \;\ldots\; 0}_{p_1\ell \text{ times}} \;,\; \underbrace{0 \;\ldots\; v_{j1} \;\ldots\; 0}_{p_1 \text{ elements}} \;,\; \underbrace{0 \;\ldots\; v_{j2} \;\ldots 0}_{} \;\;\ldots\; 0 \; v_{jT-\ell} \;\; 0 \;.\; 0]$$

i'th position

The first r.h.s. term of (4.14b.) may be written as

(4.15.) $\quad -\text{vec}\,(\hat{\Sigma}_v^{-1}\, V'VH_{110}^{-1})$

where $\hat{\Sigma}_v = \frac{1}{T} \sum_{t=1}^{T} \underline{v}_t \underline{v}_t' = \frac{1}{T} V'V$ with

$$\underset{T\times p_1}{V} = \begin{bmatrix} v_{11} & v_{p_1 1} \\ & \\ v_{1T} & v_{p_1 T} \end{bmatrix}$$

and $V\, H_{110}^{-1} = W$ is the matrix of reduced form disturbances. But

$$-\text{vec}\,(\hat{\Sigma}_v^{-1} V'W) = -(W' \otimes \hat{\Sigma}_v^{-1})\,\text{vec}\,(V') = -(W' \otimes I_{p_1})(I_T \otimes \hat{\Sigma}_v^{-1})\underline{v}$$

We can write the set of first order conditions for a maximum of the likelihood function with respect to $\underline{\eta}' = (\underline{\eta}_1', \underline{\beta}')$ after substitution of the first order conditions for Σ_v, i.e. $\hat{\Sigma}_v = \frac{1}{T} \sum_{t=1}^{T} \underline{v}_t \underline{v}_t'$,

(4.16.) $\quad \dfrac{\partial \ln L}{\partial \underline{\eta}} = Z'F'^{-1}\,(I_T \otimes \hat{\Sigma}_v^{-1})\; F^{-1}\,\underline{u} = \underline{0}$

where $Z = (Y_0,\, Y_1 - F(W' \otimes I_{p_1}),\, X_1,\, X_2)$.

As discussed by Reinsel (1977), neglecting terms which, divided by T, have zero probability limit as $T \to \infty$, we have

$$-\underset{T\to\infty}{\text{plim}}\, \frac{1}{T}\, \frac{\partial^2 \ln L}{\partial \underline{\eta}\,\partial \underline{\eta}'}\,(\hat{\Sigma}_v) = \underset{T\to\infty}{\text{plim}}\, \frac{1}{T}\; Z'F'^{-1}\,(I_T \otimes \hat{\Sigma}_v^{-1})\; F^{-1}\, Z.$$

Using a lemma by Dhrymes and Taylor (1976) the following two-step estimator for $\underline{\eta}$ has the same asymptotic distribution as the ML estimator:

$$(4.17.)\qquad \hat{\hat{\underline{\eta}}} = \hat{\underline{\eta}} - \Gamma(\hat{\underline{\eta}})^{-1} \frac{\partial \ln L}{\partial \underline{\eta}} (\hat{\underline{\eta}})$$

where $\hat{\underline{\eta}}$ is a consistent estimator of $\underline{\eta}$ such that $\sqrt{T}(\hat{\underline{\eta}} - \underline{\eta}_0)$, $\underline{\eta}_0$ being the true parameter value, has some limiting distribution and the matrix $\Gamma(\hat{\underline{\eta}})$ is such that

$$\operatorname*{plim}_{T \to \infty} \frac{1}{T} \Gamma(\hat{\underline{\eta}}) = \operatorname*{plim}_{T \to \infty} \frac{1}{T} \frac{\partial^2 \ln L}{\partial \underline{\eta} \partial \underline{\eta}'} (\underline{\eta}_0)$$

Applying (4.17.) to the present problem yields

$$(4.18.)\qquad \hat{\hat{\underline{\eta}}} = \hat{\underline{\eta}} + [\hat{Z}'\hat{F}'^{-1} (I_T \otimes \hat{\Sigma}_v^{-1}) \hat{F}^{-1} \hat{Z}]^{-1} \hat{Z}'\hat{F}'^{-1} (I_T \otimes \hat{\Sigma}_v^{-1})\hat{F}^{-1} \hat{\underline{u}}$$

where "^" denotes that the unknown quantities are evaluated at consistent parameters estimates. The first step consistent estimates can be obtained using one of the single equation estimation method proposed in section 4.1.

As we can write the system in (4.12.) as

$$(4.19.)\qquad \bar{\underline{y}} = [\underline{y} - F(W' \otimes I_{p_1}) + \underline{u} - \underline{v}] = Z\,\underline{\eta} + \underline{u} = Z\,\underline{\eta} + F\,\underline{v},$$

we may apply generalized least squares to (4.19.) after having evaluated the regressand $\bar{\underline{y}}$, the regressors Z and the disturbance covariance matrix $F'(I_T \otimes \Sigma_v)F$ at consistent parameter estimates - this is in fact one way of computing the two-step estimator in (4.18.) and it shows that the two-step estimator (4.18.) can be interpreted as a residual-adjusted estimator, a term introduced by Hatanaka (1974). Reinsel (1977) derives a slightly different estimator to which he gives an instrumental variables interpretation. It is obvious that the computation of the two-step estimator (4.18.) which, if iterated until convergence, yields the ML estimator given fixed and known initial conditions, involves the inverse of the p_1Txp_1T disturbance covariance matrix $F'(I_T \otimes \Sigma_v)F$. In the way, we have analyzed the problem, this involves the inversion of F which is a block-band triangular matrix. As shown by Palm (1977b), it only requires multiplication and addition of matrices of order p_1xp_1.

As already discussed in section 2, the approximation in (4.18.) to the second step of the Newton-Raphson algorithm, is in fact the second step of the Gauss-Newton algorithm starting from consistent parameter estimates. The large sample covariance matrix of $\hat{\hat{\underline{\eta}}}$ is consistently estimated by

(4.20.) $\hat{V}(\hat{\underline{\eta}}) = [\hat{Z}'\hat{F}'^{-1} (I_T \otimes \hat{\Sigma}_V^{-1}) \hat{F}^{-1} \hat{Z}]^{-1}$

Since $\hat{\underline{\eta}}$ will be approximately normally distributed in large samples, approximate tests of hypotheses can be constructed.

As in the discussion of TF estimation, it is important to emphasize that (4.18.) involves rather large matrices when the dimensionality of $\underline{\eta}$ is large. The situation is similar to that encountered in three-stage least-squares but here in addition to structural coefficients, there are also parameters of the MA disturbance process to estimate. As with three-stage least-squares, the estimation approach described above can be applied to subsets of the structural equations.

4.3. Single Equation Structural Estimation Reconsidered: 2 step LIML

Given that full information methods usually involve complicated computations and that the complete system is not always fully specified, we consider in this section single-equation methods from a ML point of view.

Consider a structural equation, assumed to be identified by exclusion restrictions, of the system (4.1.), say the first one

(4.21.) $Y_{(1)}\underline{\eta}_{(1)}+X_{(1)}\underline{\beta}_{(1)} = \underline{u}_1$

where $Y_{(1)} = (\underline{y}_1 \; Y_1)$ is the $T x m_{(1)}$ matrix of observations on the current endogenous variables included in the first equation, with $m_{(1)} = m_1+1$

$X_{(1)}$ is the matrix of observations on included lagged endogenous , included current and lagged exogenous variables and a column of 1's for the constant term.

$\underline{\eta}_{(1)}$ and $\underline{\beta}_{(1)}$ are vectors of the non zero structural coefficients in the first equation.

$\underline{u}_1' = (u_{11} \;\; u_{12} \;\; \ldots \;\; u_{1T})$.

We write the unrestricted reduced form for $Y_{(1)}$ as

(4.22.) $Y_{(1)} = Z\,\Pi_{(1)} + V_{(1)} = Z_1\,\Pi_{1.} + Z_0\,\Pi_{0.} + V_{(1)}$

where $\Pi'_{(1)} = (\Pi'_{1.} \vdots \Pi'_{0.})$

Postmultiplying (4.22.) by $\underline{\eta}_{(1)}$ and comparing the result with (4.21.) indicates the following restrictions:

$$(4.23.) \quad \Pi_{0.}\ \underline{\eta}_{(1)} = 0$$

$$\Pi_{1.}\ \underline{\eta}_{(1)} = \underline{\beta}_1$$

From the assumptions on the model, the rows of $V_{(1)}$ are normally distributed, with zero mean and common covariance matrix $\Omega_{(1)}$. Each row of $V_{(1)}$ can be represented as a MA of order $q_{(1)}$. In order to get a simple structure for the disturbance term covariance matrix, we vectorize the model (4.22.) as follows

$$(4.24.) \quad \text{vec}\ (Y'_{(1)}) = [Z_1 \otimes I_{m_{(1)}}]\ \text{vec}\ (\Pi'_{1.}) + [Z_0 \otimes I_{m_{(1)}}]\ \text{vec}\ (\Pi'_{0.}) + \text{vec}\ (V'_{(1)}).$$

which we write as

$$(4.25.) \quad \underline{y}_{(1)} = W_1\underline{\pi}_1 + W_0\underline{\pi}_0 + \underline{v}_{(1)}$$

and the MA representation of $\underline{v}_{(1)}$ as

$$(4.26.) \quad \underline{v}_{(1)} = F_{(1)}\ \underline{\varepsilon}_{(1)}$$

where

$$(4.27) \quad F_{(1)} = \begin{bmatrix} I_{m_{(1)}} & 0 & \cdots & \cdots & 0 \\ F_{(1)1} & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ F_{(1)q_{(1)}} & & \ddots & \ddots & 0 \\ 0 & \ddots & & \ddots & \vdots \\ 0 & 0 & F_{(1)q_{(1)}} \cdots & F_{(1)1} & I_{m_{(1)}} \end{bmatrix}$$

and $\underline{\varepsilon}_{(1)}$ is normally distributed with covariance matrix $\Omega_{(1)}$ and zero serial correlations. The likelihood function may then be written as

$$(4.28.) \quad L(\underline{y}_{(1)}, W_1, W_0, \underline{\pi}_1, \underline{\pi}_0, \Omega_{(1)}, F_{(1)}) \propto |\Omega_{(1)}|^{-T/2} \exp\ (S)$$

where $S = -\frac{1}{2}\ (\underline{y}_{(1)} - W_1\underline{\pi}_1 - W_0\underline{\pi}_0)'\ \Omega^{-1}\ (\underline{y}_{(1)} - W_1\underline{\pi}_1 - W_0\underline{\pi}_0)$

and $\quad \Omega = F'_{(1)}\ (I_T \otimes \Omega_{(1)})\ F_{(1)}$.

We define the LIML estimator in a way slightly different from the usual definition* as the estimator which maximizes (4.28.) with respect to $\underline{\pi}_1$, $\underline{\pi}_0$, $F_{(1)}$ and $\underline{\eta}_{(1)}$ subject to $\Pi_0 \underline{\eta}_{(1)} = 0$. The restrictions may also be vectorized as:

$$(4.29.) \qquad [I_m \otimes \underline{\eta}'_{(1)}]\ \underline{\pi}_0 = 0$$

The Lagrangean expression is

$$(4.30.) \qquad z = -\frac{T}{2} \log |\Omega_{(1)}| + S - \underline{\lambda}'\{[I_{m_{(1)}} \otimes \underline{\eta}'_{(1)}]\ \underline{\pi}_0\}$$

where $\underline{\lambda}$ is the $k_0 \times 1$ vector of Lagrange multipliers.
The set of first order conditions for a maximum is

$$(4.31.a.) \quad \frac{\partial z}{\partial \underline{\pi}_1} = W_1'\ \Omega^{-1}\ \underline{\varepsilon}_{(1)} = \underline{0}$$

$$(4.31.b.) \quad \frac{\partial z}{\partial \underline{\pi}_0} = W_0'\ \Omega^{-1}\ \underline{\varepsilon}_{(1)} - (I_{m_{(1)}} \otimes \underline{\eta}_{(1)})\ \underline{\lambda} = \underline{0}$$

$$(4.31.c.) \quad \frac{\partial z}{\partial \underline{\phi}} = W_2'\ \Omega^{-1}\ \underline{\varepsilon}_{(1)} = \underline{0}$$

where $\underline{\phi}' = \text{vec}\ [F_{(1)1}, F_{(1)2}, \ldots, F_{(1)q_1}]'$

$$W_2' = [\underline{\varepsilon}_{ij}^{\ell'}]\ ,\ i=1,2,\ldots,m_{(1)}\ ,\ \ell=1,2,\ldots,q_{(1)}\ ,\ j=1,2,\ldots,m_{(1)} \quad \text{and}$$

$$\underline{\varepsilon}_{ij}^{\ell'} = [\underbrace{0,0\ldots0}_{\ell\, m_{(1)} \text{ times}},\underbrace{0\ldots\varepsilon_{j1},0\ldots0}_{m_{(1)} \text{ elements}},\underbrace{0\ \ldots\varepsilon_{j2},0..0}_{},\ldots\underbrace{0\ .\ 0\ \varepsilon_{jT-\ell}0\ldots0}_{}]$$

i'th position

$$(4.31.d.) \quad \frac{\partial z}{\partial \underline{\eta}_1} = (I_{m_{(1)}} \otimes \underline{\lambda}')\ B\ \underline{\pi}_0 = 0$$

$$B = \begin{pmatrix} B_1 \\ B_2 \\ B_{m_{(1)}} \end{pmatrix} \text{ and } B_i = [I_{m_{(1)}} \otimes \underline{e}'_i]$$

$$\underline{e}_i' = [0,0..0\ 1\ 0\ ..\ 0]$$

i'th position

$$(4.31.e.) \quad \frac{\partial z}{\partial \underline{\lambda}} = [I_{m_{(1)}} \otimes \underline{\eta}'_{(1)}]\ \underline{\pi}_0 = \underline{0}$$

* In terms of asymptotic properties of the LIML estimator it does not matter whether we maximize the likelihood function concentrated with respect to $\Omega_{(1)}$ or use a consistent estimate for $\Omega_{(1)}$ in the first order conditions for a maximum of the likelihood function with respect to the remaining parameters.

We can solve (b) for $\underline{\pi}_0$, to get

$$(4.32.)\quad \underline{\pi}_0 = (W_0'\,\Omega^{-1}\,W_0)^{-1}\,[\underline{y}_1 - W_1\underline{\pi}_1 - (I_{m_{(1)}} \otimes \underline{\eta}_{(1)})\,\underline{\lambda}]$$

Substituting (4.32.) into (4.31.e.) and solving for $\underline{\lambda}$ gives

$$(4.33.)\quad \underline{\lambda} = [(I_{m_{(1)}} \otimes \underline{\eta}_{(1)}')(W_0'\,\Omega^{-1}\,W_0)(I_{m_{(1)}} \otimes \underline{\eta}_{(1)})]^{-1}[(I_{m_{(1)}} \otimes \underline{\eta}_{(1)}')$$

$$(W_0'\,\Omega^{-1}\,W_0)^{-1}\,(y_1 - W_1\underline{\pi}_1)].$$

The set of first order conditions for a maximum in (4.31.) is clearly nonlinear in the parameters. We can approximate the solution by a two-step Newton-Raphson procedure, as has been done in (2.13.), starting with consistent estimates for $\underline{\pi}_0$, $\underline{\pi}_1$, $\underline{\phi}$ and $\underline{\eta}_{(1)}$, computing $\underline{\lambda}$ from expression (4.33.) and evaluating*

$$(4.34.)\quad \hat{\hat{\underline{\theta}}} = \hat{\underline{\theta}} - \Gamma(\hat{\underline{\theta}})^{-1}\left[\frac{\partial z}{\partial \underline{\theta}}\right]_{\underline{\theta}=\hat{\underline{\theta}}}$$

where $\underline{\theta} = (\underline{\pi}_1', \underline{\phi}', \underline{\pi}_0', \underline{\eta}_{(1)}', \underline{\lambda}')$, $\hat{\underline{\theta}}$ is a consistent estimate of $\underline{\theta}$ satisfying the requirement in (2.13.) and

$$(4.35.)\quad \Gamma(\underline{\theta}) = \begin{bmatrix} W_1'\Omega^{-1}W_1 & W_1'\Omega^{-1}W_2 & W_1'\Omega^{-1}W_0 & 0 & 0 \\ W_2'\Omega^{-1}W_1 & W_2'\Omega^{-1}W_2 & W_2'\Omega^{-1}W_0 & 0 & 0 \\ W_0'\Omega^{-1}W_1 & W_0'\Omega^{-1}W_2 & W_0'\Omega^{-1}W_0 & -B'\underline{\lambda} & -(I_{m_{(1)}} \otimes \underline{\eta}_{(1)}) \\ 0 & 0 & -\underline{\lambda}'B & 0 & B'B\underline{\pi}_0 \\ 0 & 0 & -(I_{m_{(1)}} \otimes \underline{\eta}'_{(1)}) & \underline{\pi}_0'B'B & 0 \end{bmatrix}$$

The probability limit of the matrix $\frac{1}{T}\Gamma(\hat{\underline{\theta}})$ in (4.35.) usually is the matrix $\frac{1}{T}\frac{\partial^2 z}{\partial\underline{\theta}\partial\underline{\theta}'}(\underline{\theta}_0)$ where $\underline{\theta}_0$ is the vector of true parameter values of $\underline{\theta}$. Of course, one can iterate the expression (4.34.) to get the exact solution of the first order conditions for a maximum of the likelihood function, which is the limited information ML estimator given fixed and known initial conditions. In terms of asymptotic efficiency, it is not necessary to continue the iteration after the second step.

*The unknown elements of $\Omega_{(1)}$ will also be replaced by consistent estimates.

V. Some concluding remarks

1. In this paper, we have presented several estimators for the three forms of a dynamic SEM with moving average disturbances, and we have discussed their asymptotic properties. The results essentially rely upon

a) the asymptotic properties of the ML estimator of the parameters of dynamic models and

b) upon a result given by Fisher (1925), Kendall and Stuart (1961), Rotherberg and Leenders (1964) and later by Dhrymes and Taylor (1976) concerning the asymptotic porperties of a two-step iteration of the first order conditions for a maximum of the likelihood function.

Of course, the starting values for the iteration and the matrix Γ approximating the matrix of second-order derivation of the log-likelihood function have to satisfy some conditions (see e.g. (2.13.)), which we give in the text, but which we do not verify explicitly for the estimation problems considered. It ought to be clear that the requirements such as stated in (2.13.) have to be checked in practical situations.

2. Computation of the estimators presented above generally involves operations on large matrices. For ex., in each case one has to compute the inverse of the covariance matrix of a vector-MA proces.
The estimation methods presented here open an immense field of application for good numerical matrix inversion procedures exploiting the special features of the covariance matrix of an MA process.

3. Despite the fact that the field of application of the methods presented is probably limited to small models, the results of the paper clarify a number of questions concerning the asymptotic properties of estimators for dynamic and static models. For example, if the disturbances of the TF system in (3.6) are not correlated, i.e. $K_h=0$, $h=1,2,\dots,q$, then the two-step estimator given in (3.10.) specialises to Zellner's estimator for seemingly unrelated regressions.

As a second example, assume that $H_{11}(L)$ in (3.3a.) is an unimodular matrix, i.e. $|H_{11}(L)|=$ constant, then the expression given in (3.10.) specializes to an expression with $\hat{Z}' = (X_0', X'\hat{Z}_2')$ and the covariance matrix of the estimator in (3.10.) will be asymptotically a block-diagonal matrix as $\underset{T\to\infty}{\text{plim}} \frac{1}{T} [(X_0'X')\hat{Z}_2] = 0$ under suitable conditions. Therefore it will be sufficient to have consistent estimates of $\underline{\lambda}_2$ to efficiently estimate $(\underline{c}_1', \underline{\delta}')$ in (3.7.). A similar result has been established by Amemiya (1973).

4. It is to be expected that the estimation results can - at least - for samples of the size encountered in applied work - be improved by using two-step estimators approximating the first order conditions for a maximum

of the exact likelihood function. One step in the direction of using the exact likelihood function is to "backforecast" the values of the initial conditions for FE, TF or structural equation systems. This aspect however deserves additional work.

5. The discussion has been in terms of large sample properties of the estimators for dynamic models. Small sample properties of the estimators have to be investigated yet. However the Monte Carlo results obtained by Nelson (1976) justify some optimism about improving in small samples the efficiency of the estimation results by computing the second-step of iteration.

REFERENCES

Amemiya, T. (1973): "Generalized Least Squares with Estimated Autocovariance Matrix", Econometrica, 14, 723-732.

Aigner, D.J. (1971): "A Compendium on Estimation of the Autoregressive-Moving Average Model from Time Series", International Economic Review, 12, 348-371.

Akaike, H. (1973): "Maximum Likelihood Identification of Gaussian Autoregressive-Moving Average Models", Biometrika, 60, 225-265.

Almon, S. (1965): "The Distributed Lag Between Capital Appropriations and Expenditures", Econometrica, 33, 178-196.

Anderson, T.W. (1975): "Maximum Likelihood Estimation of Parameters of Autoregressive Processes with Moving Average Residuals and other Covariance Matrices with Linear Structure", The Annals of Statistics, 3, 1283-1304.

Ansley, C.F., Spivey, W.A., and W.J. Wroblenski (1977): "On the Structure of Moving Average Processes", Journal of Econometrics, 6, 121 - 134.

Åström, K.J., and T. Bohlin (1966): "Numerical Identification of Linear Dynamic Systems from Normal Operating Records", in Ph. H. Hammond (Ed.): Theory of Self-Adaptive Control Systems, New York: Plenum Press, 1966.

Box, G.E.P., and G.M. Jenkins (1970): Time Series Analysis, Forecasting and and Control, San Francisco: Holden-Day, Inc.

Byron, R.P. (1973): "The Computation of Maximum Likelihood Estimates for Linear Simultaneous Systems with Moving Average Disturbances", mimeo, Australian National University.

Chow, G.C., and R.C. Fair (1973): "Maximum Likelihood Estimation of Linear Equation Systems with Autoregressive Residuals", Annals of Economic and Social Measurement, 2, 17-28.

Deistler, M. (1975): "Z-Transform and Identification of Linear Econometric Models with Autocorrelated Errors", Metrika, 22, 13-25.

Deistler, M. (1976): "The Identifiability of Linear Econometric Models with Autocorrelated Errors", International Economic Review, 17, 26-46.

Dhrymes, P.J. (1971): Distributed Lags, San Francisco: Holden-Day, Inc.

Dhrymes, P.J., and J.B. Taylor (1976): "On an Efficient Two-Step Estimator for Dynamic Simultaneous Equations Models with Autoregressive Errors", International Economic Review, 17, 362-376.

Durbin, J. (1959): "Efficient Estimation of Parameters in Moving-Average Models", Biometrika, 46, 306-316.

Espasa, A. (1975): The Spectral Estimation of the Simultaneous Equations Model and its Relation with Time Domain Methods", mimeo, London School of Economics.

Espasa, A., and J.D. Sargan (1975): "The Spectral Estimation of Simultaneous Equations System with Lagged Endogenous Variables", mimeo, paper presented at the Third World Econometric Congress.

Feller, W. (1966): An Introduction to Probability Theory and its Applications, vol. 2, New York, John Wiley & Sons, Inc., 2nd edit.

Fisher, R.A. (1925): Statistical Methods for Research Workers, Edinburgh, Oliver and Boyd.

Goldfeld , S.M. and R.E. Quandt (1972): Non-Linear Methods in Econometrics, Amsterdam, North-Holland Publ. Co.

Granger, C.W., and M.J. Morris (1976): "Time Series Modelling and Interpretation", Journal of the Royal Statistical Society, A, 139, 246-257.

Hannan, E.J. (1969): "The Identification of Vector Mixed Autoregressive-Moving Average Systems", Biometrika, 56, 223-225.

Hannan, E.J. (1971): "The Identification Problem for Multiple Equation Systems with MA Errors", Econometrica, 39, 751-765.

Hannan, E.J. and R.D. Terrell (1973): "Multiple Equation Systems with Stationary Errors", Econometrica, 41, 299-320.

Hannan, E.J. (1975): "The Estimation of ARMA Models", The Annals of Statistics, 3, 975-981.

Hatanaka, M. (1974): "An Efficient Two-Step Estimator for the Dynamic Adjustment Model with Autoregressive Errors", Journal of Econometrics, 2, 199 - 220.

Hatanaka, M. (1975): "On Global Identification of the Dynamic Simultaneous Equations Model with Stationary Disturbances", International Economic Review, 16, 545-554.

Hatanaka, M. (1976): "Several Efficient Two-Step Estimators for the Dynamic Simultaneous Equations Model with Autoregressive Disturbances", Journal of Econometrics, 4, 189-204.

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

Kang, K.M. (1975): "A Comparison of Estimators for Moving Average Processes", unpublished paper, Australian Bureau of Statistics.

Kendall, M.G., and A. Stuart (1961): Advanced Theory of Statistics, Vol. 2, London, Griffin and Co.

Kmenta, J., and R.F. Gilbert (1968): "Small Sample Properties of Alternative Estimates of Seemingly Unrelated Regressions", Journal of the American Statistical Association, 63, 1180-1200.

Maddala, G.S. (1971): "Generalized Least Squares with Estimated Variance Covariance Matrix", Econometrica, 39, 23-33.

Nelson, C.R. (1976): "Gains in Efficiency from Joint Estimation of Systems of Autoregressive-Moving Average Processes", Journal of Econometrics, 4, 331 - 348.

Nicholls, D.F. (1976): "The Efficient Estimation of Vector Linear Time Series Models", Biometrika, 63, 381-390.

Nicholls, D.F., Pagan, A.R., and R.D. Terrell (1975): "The Estimation and Use of Models with Moving Average Disturbance Terms: A Survey", International Economic Review, 16, 113-134.

Osborn, D.R. (1976): "Maximum Likelihood Estimation of Moving Average Processes", Annals of Economic and Social Measurement, 5, 75-87.

Palm, F. (1977a): "On Univariate Time Series Methods and Simultaneous Equation Models", Journal of Econometrics, 5, 379-388.

Palm, F. (1977b): "On Efficient Estimation of the Final Equation Form of a Linear Multiple Time Series Process", Cahiers du Centre d'Etudes de Recherche Opérationnelle,19, 297-308.

Pesaran, M.H. (1973): " Exact Maximum Likelihood Estimation of a Regression with a First-Order Moving Average Error", Review of Economic Studies, 41, 529-536.

Phillips, A.W. (1966): "Estimation of Systems of Difference Equations with Moving Average Disturbances",mimeo, London School of Economics.

Pierce, D.A. (1972): "Least Squares Estimation in Dynamic-Disturbance Time Series Models", Biometrika, 59, 73-78.

Quenouille, M.H. (1968): The Analysis of Multiple Time Series, London: C. Griffin and Co., 2nd edition.

Rothenberg, T.J., and C.T. Leenders (1964): "Efficient Estimation of Simultaneous Equation System", Econometrica, 32, 57-76.

Reinsel, G. (1976): "Maximum Likelihood Estimation of Vector Autoregressive Moving-Average Models", mimeo, Department of Statistics, Carnegy-Mellon University.

Reinsel, G. (1977): "FIML Estimation of the Dynamic Simultaneous Equations Model with ARMA Disturbances", mimeo, University of Wisconsin.

Shiller, R.J. (1973): "A Distributed Lag Estimator Derived from Smoothness Priors", Econometrica, 41, 775-788.

Swamy, P.A.V.B., and P.N. Rappoport (1978), "Relative Efficiencies of Some Simple Bayes Estimators of Coefficients in a Dynamic Equation with Serially Correlated Errors - II", Journal of Econometrics, 7, 245 - 258.

Wall, K.D. (1976): "FIML Estimation of Rational Distributed Lag Structural Models", Annals of Economic and Social Measurement, 5, 53-62.

Wilson, G.T. (1973): "The Estimation of Parameters in Multivariate Time Series Models", Journal of the Royal Statistical Society, B, 76-85.

Zellner, A. (1971): "Bayesian and Non-Bayesian Analysis of the Log-Normal Distribution and Log-Normal Regression", Journal of the American Statistical Association, 66, 327-330.

Zellner, A. (1971): An Introduction to Bayesian Inference in Econometrics, New York: John Wiley and Sons.

Zellner, A., and W. Vandaele (1974): "Bayes-Stein Estimators for k-Means, Regressions and Simultaneous Equation Models", in Fienberg, S.E., and A. Zellner (eds.): Studies in Bayesian Econometrics and Statistics, North-Holland.

Zellner, A., and F. Palm (1974): "Time Series Analysis and Simultaneous Equation Econometric Models", Journal of Econometrics, 2, 17-54.

Zellner, A., and F. Palm (1975): "Time Series and Structural Analysis of Monetary Models of the U.S. Economy", Sankhyā: The Indian Journal of Statistics, C, 37, 12-56.