Evaluating Forecast Uncertainty in Econometric Models: The Effect of Alternative Estimators of Maximum Likelihood Covariance Matrix

Calzolari, Giorgio and Panattoni, Lorenzo

IBM Scientific Center, Pisa, Italy.

8 July 1984

EVALUATING FORECAST UNCERTAINTY IN ECONOMETRIC MODELS: THE EFFECT OF ALTERNATIVE ESTIMATORS OF MAXIMUM LIKELIHOOD COVARIANCE MATRIX

by Giorgio CALZOLARI and Lorenzo PANATTONI

Centro Scientifico IBM - Pisa

Most of the methods proposed in the literature for evaluating forecast uncertainty in econometric models need an estimate of the structural coefficients covariance matrix among input data. When estimation is performed with full information maximum likelihood, alternative estimators of such a covariance matrix (Hessian, outer product, generalized least squares type matrix, quasi maximum likelihood type matrix), although asymptotically equivalent, often produce large differences in practical applications. Experimental results will be given for some econometric models well known in the literature, both with historical data and with data generated by Monte Carlo.

1. INTRODUCTION

Econometric models are often used to forecast future outcomes of economic variables. The forecast obviously depends on the estimated values of the structural coefficients and the fact that asymptotically equivalent estimators may have quite different distributions for the sample sizes occurring in practice is so well known that no model builder feels anymore in trouble when noticing that coefficients obtained by applying different estimation algorithms to his model may look quite different one from another.

However full information maximum likelihood is commonly considered one of the most appealing among all the estimators from the theoretical point of view and recently some algorithms have been proposed for a computationally efficient use of this method even for medium size nonlinear models (see for example Parke, 1982, or Calzolari and Panattoni, 1983), in order to make it of practical interest to the model builders. It can therefore be interesting to analyze the case in which the model builder is in the lucky situation of applying succesfully the FIML technique to his model before using the model for forecasting purposes. Unless particular situations arise, he will get a univocal set of estimated coefficients and a univocal set of forecasts for the endogenous variables of his model.

But the model user is interested also in knowing the reliability of his forecast and from this point of view the problem looses its unambiguousness. In fact in most cases an evaluation of this reliability will be based, among other things, also on the knowledge of an estimate of the structural coefficients covariance matrix as supplied by the computer program used to get the estimate of the coefficients, and several possibilities, leading to different choices of this matrix, are available in the literature, for example:

- a generalized least squares type matrix, if maximization is performed by iterating some suitable generalized least squares or instrumental variables algorithms, as in Dagenais (1978), or Hausman (1974);
- the inverse of the Hessian of the concentrated log-likelihood, if maximization is based on some Newton-like algorithm, like that of Eisenpress and Greenstadt (1966);
- the inverse of the outer product of the first derivatives of the log-likelihoods, if the algorithm takes advantage of the suggestions in Berndt et al. (1974).

Although all these matrices are asymptotically equivalent, the differences arising from their use for practical sample periods are often so big that they can lead the model user to opposite conclusions and decisions when testing hypotheses on the structural coefficients themselves (see Calzolari and Panattoni, 1983), or when judging the reliability of the model in forecasting. The aim of this paper is just to get a deeper insight into this problem by analyzing the differences in the evaluation of the forecast's reliability, arising from the use of different estimates of the FIML covariance matrix, when FIML is applied to some small-medium sized econometric models available in the literature.

2. NOTATIONS

Let the simultaneous equation model be represented as

(1) $f_i(y_t, x_t, a_i) = u_{it}$ $\qquad i=1,2,\ldots,m;\ t=1,2,\ldots,T$

where y_t is the m×1 vector of endogenous variables at time t, x_t is the vector of exogenous variables at time t and a_i is the vector of unknown structural coefficients in the i-th equation. The m×1 vector of random error terms at time t, $u_t=(u_{1t}, u_{2t}, \ldots, u_{mt})'$, is assumed to be

independently and identically distributed as $N(0,\Sigma)$, with Σ completely unknown, apart from being symmetric and positive definite. The complete n×1 vector of unknown structural coefficients of the system will be indicated as $a=(a_1', a_2', \ldots, a_m')'$.

If the model is linear both in the variables and in the coefficients, its structural form can be represented as

(2) $Ay_t + Bx_t = u_t$

where A is the m×m matrix of structural coefficients of endogenous variables, and B is the matrix of coefficients of exogenous variables. A and B are sparse matrices, if the model's equations are overidentified (as it usually happens in practice), whose unknown terms are the elements of the vector a, in the notation (1).

It is usually assumed that a simultaneous equation system like (1) uniquely defines the values of the elements of y_t once values for the coefficients, the predetermined variables and the disturbance terms are given (at least in some range); in the case of the linear model (2) this is equivalent to assuming nonsingularity of the matrix A. This means that the structural form equations (1) implicitly define a system of reduced form equations

(3) $y_t = \gamma(x_t, a, u_t)$

where the vector of functional operators γ is generally unknown in case of nonlinear models.

If the model is linear, equation (3) simply becomes

(4) $y_t = \Pi x_t + v_t;\qquad \Pi = -A^{-1}B;\qquad v_t = A^{-1}u_t.$

3. EX-POST FORECASTS AND FORECAST ERRORS

Let $\hat{a}$ be the full information maximum likelihood estimate of the vector a, obtained by using the data for y_t and x_t in the sample period t=1,2,...,T, and let us use the model to produce a forecast in a forecast period h not belonging to 1,2,...,T. We shall deal only with ex-post forecasts, that is assuming exact knowledge of all the predetermined variables in the forecast period x_h. Under the assumption of serial independence, the disturbances at time h, u_h, are independent of the disturbance terms in the sample period and, therefore, u_h and $\hat{a}$ are two independent random variables.

The usual forecast supplied by the model is obtained by inserting, in the structural form equations (1), the estimated vector $\hat{a}$ and the values of the predetermined variables x_h (supposed exact, for the purposes of this paper, as already observed), dropping the disturbance term and solving the resulting system

(5) $$f_i(\hat{y}_h, x_h, \hat{a}_i) = 0 \qquad i=1,2,\dots,m$$

by means of some numerical method. In terms of the (unknown) reduced form, this means that the vector of forecasts at time h can be represented as

(6) $$\hat{y}_h = y(x_h, \hat{a}, 0).$$

The vector of the true values of endogenous variables in the forecast period can be represented, using the unknown reduced form, as

(7) $$y_h = y(x_h, a, u_h).$$

The vector of forecast errors is the difference between $\hat{y}_h$ and y_h. It is now convenient to introduce an auxiliary vector, $\bar{y}_h$, defined as the

vector of forecasts that would be produced by the model if the structural coefficients were known with certainty; in other words $\bar{y}_h$ is the solution of the model free of errors at time h.

(8) $$f_i(\bar{y}_h, x_h, a_i) = 0 \qquad i=1,2,\dots,m$$

that is, in terms of reduced form,

(9) $$\bar{y}_h = y(x_h, a, 0).$$

Returning to the vector of forecast errors, we now have

(10) $$\begin{aligned}\hat{y}_h - y_h &= [\hat{y}_h - \bar{y}_h] + [\bar{y}_h - y_h] \\ &= [y(x_h, \hat{a}, 0) - y(x_h, a, 0)] + [y(x_h, a, 0) - y(x_h, a, u_h)].\end{aligned}$$

In the case of linear model, equation (10) assumes the well known form

(11) $$\begin{aligned}\hat{y}_h - y_h &= [\hat{\Pi}x_h - \Pi x_h] + [\Pi x_h - (\Pi x_h + v_h)] \\ &= [\hat{\Pi} - \Pi]x_h - A^{-1}u_h.\end{aligned}$$

In both cases, the vector of forecast errors is the sum of two random vectors: the first is a function of several variables, among which only the vector of estimated coefficients, $\hat{a}$, is random; the second is also a function of several variables, among which only the vector of structural disturbances, u_h, is random. Since, by assumption, $\hat{a}$ and u_h are independent, so also are the two components of the vector of forecast errors.

Therefore the two components can be separately analyzed and, in particular, an estimate of the variances of the forecast errors can be obtained by summing the estimated variances of the two components.

What we have stated above is not exactly true if lagged endogenous variables are present among the predetermined variables; in this case, in fact, the two terms of the sum are both functions of the (random) lagged endogenous variables. The above considerations, however, still hold conditional on a given value of the lagged endogenous variables (the historical values, in case of one-period forecasts).

3.1 The effect of the random error terms

To analyze the component $(\bar{y}_h - y_h)$, which is a function of the random structural disturbances, stochastic simulation is usually proposed as the basic computational method. By means of replicated solutions of the model, each time introducing a vector of pseudo-random disturbances in place of u_h, it is possible to compute approximate values of the conditional means and variances of the elements of $(\bar{y}_h - y_h)$. The approximation improves, usually, as the number of replications increases, and improves even more if some suitable variance reduction algorithm is adopted, like the algorithm based on control variates proposed in Calzolari and Sterbenz (1983). If finite moments of the first two orders exist, a very high number of replications would lead, in practice, to the exact values of means and variances, if the parameters of the model (the vector a and the covariance matrix of the structural disturbances) were known with certainty. As, however, we are assuming to have only FIML estimates of these parameters, stochastic simulation will lead to an estimate of the means and variances of the elements of $(\bar{y}_h - y_h)$.

Of course, if the model is linear, the mean of this component is zero and the covariance matrix of its elements is

(12) $A^{-1}\Sigma A'^{-1}$ and $\hat{A}^{-1}\hat{\Sigma}\hat{A}'^{-1}$ its FIML estimate.

As already mentioned, this method is accepted and used practically in all the works on this topic. Possible slight variants are possible, depending on the computation of the matrix $\hat{\Sigma}$. It is in all cases computed from the [illegible] form residuals, usually as

(13) $\hat{\Sigma} = T^{-1}\sum_t \hat{u}_t \hat{u}_t'$

but it might also take into account the problem of degrees of freedom, analogously to the linear regression model (e.g. Klein, 1969). It might also not be computed explicitly, as is the case of McCarthy's algorithm for stochastic simulation. In practical applications, however, these differences are rather small and will not be discussed further here.

3.2. The effect of errors in estimated coefficients

The rest of the paper will deal only with the other component of the forecast error, for which the literature presents methods that differ from one another both computationally and conceptually. These methods are of three types:

1) Full analytical methods: they were originally designed for linear systems (e.g. Goldberger et al., 1961) but even in case of models containing nonlinearities these methods can be applied to solve a good deal of the problem (Calzolari, 1981).
2) Mixed methods, partially analytical and partially based on numerical simulation procedures (analytic simulation), conceptually equivalent to the full analytical methods, that allow for a considerable reduction of computational complexity and are suitable for application even to medium-large size models (Bianchi and Calzolari, 1980).
3) Monte Carlo methods: estimates of the variances are computed from sample variances of replicated simulation experiments, after pseudo-random errors have been inserted into model's coefficients, directly (Fair, 1980), or through a process of simulation and re-estimation of the whole model (Schink, 1971).

Comparisons of the methods can be found in Bianchi and Calzolari (1982). In this paper we are going to use only the full analytical method [illegible] numerical [illegible] called analytic simulation, see Bianchi and

Calzolari, 1980).

The method relies on the property, well known in large sample theory, that asymptotic normality of sample statistics can be maintained through transformations, even nonlinear, provided they are continuously differentiable. If we assume that, as T increases, asymptotically

(14) $$\sqrt{T}(\hat{a} - a) \sim N(0, \Psi)$$

(and $\hat{\Psi}$ is a consistent estimate of Ψ) then, asymptotically,

(15) $$\sqrt{T}(\hat{y}_h - \bar{y}_h) = \sqrt{T}[y(x_h, \hat{a}, 0) - y(x_h, a, 0)] \sim N(0, J_h \Psi J_h')$$

where J_h is the m×n matrix of first order partial derivatives of the elements of y with respect to the elements of a, computed at the point $(x_h, a, 0)$.

If the computation is performed at the point $(x_h, \hat{a}, 0)$ and $\hat{\Psi}$ is used in equation (15), then $\hat{J}_h \hat{\Psi} \hat{J}_h'$ is a consistent estimator of $J_h \Psi J_h'$: the division by the sample period length, T, leads to the result we are looking for, the estimate of the covariance matrix of a multinormal distribution which approximates the small sample distribution of the random vector $(\hat{y}_h - \bar{y}_h)$.

Continuity and differentiability of the elements of the unknown vector of reduced form functional operators y is ensured by the implicit function theorem, which also provides the means for a full analytical computation of the partial derivatives

(16) $$\partial y/\partial a' = -(\partial f/\partial y')^{-1}(\partial f/\partial a')$$

since the structural form operators vector $f=(f_1, f_2, \dots, f_m)'$ is known (of course, some numerical solution method must be first used to get the deterministic solution of the model at time h).

For medium or large scale models it can be more convenient to perform

the computation of the above derivatives with finite differences (analytic simulation).

If the model is linear, recalling equation (4) and making use of the formula proposed in Nissen (1968), the above method can be made more explicit as

(17) $$\sqrt{T}(\hat{y}_h - \bar{y}_h) = \sqrt{T}(\hat{\Pi} - \Pi) x_h = \sqrt{T} \operatorname{vec}[I (\hat{\Pi} - \Pi) x_h] = \sqrt{T}(x_h' \otimes I) \operatorname{vec}(\hat{\Pi} - \Pi)$$

where I is the m×m unit matrix.

Equation (17) represents a linear combination of the elements of $(\hat{\Pi}-\Pi)$ with fixed coefficients, so that the asymptotic covariance matrix of $\sqrt{T}(\hat{y}_h - \bar{y}_h)$ can be computed with no difficulty as soon as the asymptotic covariance matrix of $\sqrt{T}\operatorname{vec}(\hat{\Pi}-\Pi)$ has been computed, and this can be done with the methods proposed by Goldberger et al. (1961). A method for obtaining a simple explicit representation of this component of the forecast error as a function of the structural coefficients is given in Calzolari (1981).

4. FULL INFORMATION MAXIMUM LIKELIHOOD

The log-likelihood of the t-th observation can be expressed as

(18) $$L_t = -1/2 \log|\Sigma| + \log|\partial f_t/\partial y_t'| - 1/2\, f_t' \Sigma^{-1} f_t$$

where $f_t = (f_{1t}, f_{2t}, \dots, f_{mt})' = u_t$ and the Jacobian determinant $|\partial f_t/\partial y_t'|$ is taken in absolute value. The log-likelihood of the whole sample is

(19) $$L_T = \sum_t L_t ;$$

it will be hereinafter referred to as the unconcentrated log-likelihood.

Maximization of (19) with respect to the unknown parameters provides the full information maximum likelihood (FIML) estimates of the structural form coefficients. We shall suppose that no problems of multiple maxima arise, so that for each model, given the sample period, the FIML estimates of the coefficients, a, are univocally obtainable by means of one of the available algorithms (e.g. Eisenpress and Greenstadt, 1966, Berndt et al., 1974, Hausman, 1974, Dagenais, 1978, Parke, 1982, etc.)

We are now using the model with FIML estimated coefficients to produce a forecast at time h, outside the sample period. Also the forecast, produced as usual with deterministic simulation is univocally obtainable. We are now interested in associating the forecast of each endogenous variable with an indicator of uncertainty, like a standard error.

As discussed in section 3, we need among other things an estimate of the asymptotic covariance matrix of the structural coefficients. Several alternative estimators are available for this matrix. Before discussing them, some other symbols and formulae need to be introduced.

5. FIRST AND SECOND DERIVATIVES OF THE LOG-LIKELIHOODS

We define, for the i-th equation, $g_{it}=\partial f_{it}/\partial a_i$, which is a column vector with the same length as a_i; we define also, for any i and j, the matrix $g_{ijt}=\partial^2 f_{it}/\partial a_i \partial a_j'$. If $i \neq j$, g_{ijt} is zero; it is zero also for $i=j$ if the model is linear in the coefficients (even if nonlinear in the variables). We note, now, that g_{it} and g_{ijt} may be regarded as functions of u_t, x_t and a, under the standard assumption of a one-to-one correspondence between u_t and y_t. Differentiating with respect to the coefficients of the i-th equation we get

(20) $$\partial L_t/\partial a_i = \partial g_{it}/\partial u_{it} - g_{it} f_t' \sigma^i$$

(21) $$\partial L_t/\partial(\Sigma^{-1}) = 1/2\ \Sigma - 1/2\ f_t f_t'$$

where use has been made of $\partial g_{it}/\partial u_{it}=(\partial g_{it}/\partial y_t')(\partial f_t/\partial y_t')_i^{-1}$, and σ^i represents the i-th column of Σ^{-1}, and no restriction has been placed on Σ. Considering that Σ^{-1} is symmetric, differentiating with respect to its i,j-th term we get

(22) $$\partial L_t/\partial \sigma^{ij} = 1/2\ \sigma_{ij} - 1/2\ f_{it} f_{jt} \qquad (\times 2,\ \text{if } i \neq j).$$

Using $\partial g_{ijt}/\partial u_{it}=(\partial g_{ijt}/\partial y_t')(\partial f_t/\partial y_t')_i^{-1}$, further differentiation of (20) gives

(23) $$\partial^2 L_t/\partial a_i \partial a_j' = \partial g_{ijt}/\partial u_{it} - (\partial g_{it}/\partial u_{jt})(\partial g_{jt}'/\partial u_{it}) - g_{ijt} f_t' \sigma^i - \sigma^{ij} g_{it} g_{jt}'$$

(24) $$\partial^2 L_t/\partial \sigma^{ii} \partial a_i = -\ g_{it} f_{it}$$

(25) $$\partial^2 L_t/\partial \sigma^{ij} \partial a_i = -\ g_{it} f_{jt}$$

(26) $$\partial^2 L_t/\partial \sigma^{ij} \partial a_j = -\ g_{jt} f_{it}$$

(27) $$\partial^2 L_t/\partial \sigma^{ij} \partial a_r = 0 \qquad \text{if } r \neq i \text{ and } r \neq j$$

(28) $$\partial^2 L_t/\partial \sigma^{ij} \partial \sigma^{rr} = -\ 1/2\ \sigma_{ir} \sigma_{rj} \qquad (\times 2 \text{ if } i \neq j)$$

(29) $$\partial^2 L_t/\partial \sigma^{ij} \partial \sigma^{rs} = -\ 1/2\ \sigma_{ir} \sigma_{sj} - 1/2\ \sigma_{is} \sigma_{rj} \qquad \text{if } r \neq s\ (\times 2 \text{ if } i \neq j).$$

Under standard assumptions, by equating to zero the first order derivatives of the unconcentrated log-likelihood with respect to Σ^{-1} (21), and substituting back into (18) and (19), we get the concentrated log-likelihood function

(30) $$\ell_T = \sum_t \log|\partial f_t/\partial y_t'| - 1/2\ \log|T^{-1} \sum_t f_t f_t'|$$

Differentiating ℓ_T with respect to the coefficients of the i-th equation, we have

$$(31)\qquad \partial \ell_T/\partial a_i = \sum_t \partial g_{it}/\partial u_{it} - T(\sum_t g_{it} f_t')(\sum_t f_t f_t')_i^{-1}$$

which is equal to $\partial L_T/\partial a_i$, obtained from summing (20) over time, provided that covariance parameters are replaced with their FIML estimates.

Further differentiation of (31), with respect to the structural coefficients of equation j, gives the i,j-th block of the Hessian matrix of the concentrated log-likelihood

$$(32)\qquad -\partial^2 \ell_T/\partial a_i \partial a_j' = -\sum_t \partial g_{ijt}/\partial u_{it} + T(\sum_t g_{ijt} f_t')(\sum_t f_t f_t')_i^{-1}$$
$$+ [\sum_t (\partial g_{it}/\partial u_{jt})(\partial g_{jt}'/\partial u_{it})] - T(\sum_t f_t f_t')_{ij}^{-1}(\sum_t g_{it} g_{jt}')$$
$$- T(\sum_t g_{it} f_t')(\sum_t f_t f_t')_i^{-1}(\sum_t f_t f_t')_i^{-1}(\sum_t f_t g_{jt}')$$
$$- T(\sum_t f_t f_t')_{ij}^{-1}(\sum_t g_{it} f_t')(\sum_t f_t f_t')^{-1}(\sum_t f_t g_{jt}').$$

For models which are linear in the coefficients (even if nonlinear in the variables), g_{ijt} and its derivatives are zero, so that the first and third term on the right hand side of equation (23) and the first two terms on the right hand side of equation (32) vanish. Moreover, g_{it} is nothing but the vector of values, at time t, of the explanatory variables of the i-th equation. Therefore, the numerical evaluation of all the above equations requires only one order of differentiation, that is the computation of derivatives of the explanatory endogenous variables in the i-th and j-th equations with respect to the error terms of the same equations; furthermore, since $\partial g_{it}/\partial u_{st} = (\partial g_{it}/\partial y_t')(\partial f_t/\partial y_t')_i^{-1}$, this differentiation could even be performed analytically without any particular difficulty. The use of equations (23-29) and (32) for the computation of the Hessian matrices (unconcentrated and concentrated, respectively) is therefore a sufficiently manageable matter even for medium-large models. As far as our computational experience is concerned, their use with numerical

calculation of the first derivatives of the g_{it}'s usually ensured quite accurate results, while the rough second order numerical differentiation of L_T and ℓ_T (which, on their turn, involve a further order of differentiation to calculate the Jacobian determinant) is well known to produce inaccurate results at higher computational costs (see, for example, Eisenpress and Greenstadt, 1966, p.860 and also the discussion in Parke, 1982, p.94 on the difficulty of obtaining a positive definite matrix from calculating the Hessian with numerical differentiation).

The formulae given above can be used to build most of the matrices used in this study, that is all the matrices based on the Hessian of the concentrated likelihood, the Hessian of the unconcentrated likelihood, and the outer product of the first derivatives of the likelihoods. We still need to introduce one more matrix, and a simple way of doing it is to follow Amemiya's (1977) instrumental variables approach.

Introducing the T×m matrix F, whose t,i-th element is $f_i(y_t,x_t,a_i)=u_{it}$, and the matrix G_i, whose t-th row is g_{it}', then the vector of first derivatives (31) can be rewritten as

$$(33)\qquad \partial \ell_T/\partial a_i = [T^{-1}\sum_t (\partial g_{it}/\partial u_t')F' - G_i'] F(T^{-1}F'F)_i^{-1}.$$

We define now,

$$(34)\qquad \hat{G}_i = G_i - T^{-1}F\sum_t (\partial g_{it}'/\partial u_t)$$

and build the block diagonal matrices G and $\hat{G}$, whose m diagonal blocks are G_i and $\hat{G}_i$, respectively. Moreover, evaluating all terms at $\hat{a}$, we have

$$(35)\qquad T^{-1}\hat{F}'\hat{F} = \hat{\Sigma}.$$

To compute FIML estimates, we must solve the nonlinear system obtained by equating to zero the gradient of the concentrated likelihood (31) [illegible] (33) [illegible], and combining all equations for i=1,2,...,m, we get

(36) $\tilde{G}'(\tilde{\Sigma}^{-1}\otimes I)\,\mathrm{vec}\tilde{F} = 0$

where the left hand side is a compact and computationally simple expression of the gradient of the concentrated log-likelihood.

A Taylor expansion of the gradient would give the usual Newton's iterative procedure. An alternative procedure to get the maximum likelihood estimate of a is obtained from a Taylor expansion of $\mathrm{vec}\tilde{F}$ as a function of the coefficients vector, a. The simple iterative method which results is

(37) $a^{k} = a^{k-1} - [\tilde{G}'(\tilde{\Sigma}^{-1}\otimes I)\tilde{G}]^{-1}\tilde{G}'(\tilde{\Sigma}^{-1}\otimes I)\,\mathrm{vec}\tilde{F}$

A more convenient iterative method is obtained if the square matrix which appears in brackets on the right hand side of (37) is replaced by the matrix

(38) $\tilde{R} = [\tilde{G}'(\tilde{\Sigma}^{-1}\otimes I)\tilde{G}]$

which has the advantage of being symmetric and positive definite, and is asymptotically equivalent to the previous one.

A further simplification can be introduced into the above formulae if the model is linear both in the variables and in the coefficients. In this case, in fact, $\partial g_{it}/\partial w_{it}$ is no more time varying; if the model is

(39) $Ay_t - Bx_t = u_t$

then the vector $\partial g_{it}/\partial w_{it}$, for any t, is made up of zeros (corresponding to the exogenous components of g_{it}) and of elements of A^{-1} (corresponding to the endogenous components of g_{it}). The endogenous elements of $\tilde{G}$ would simply be the values of the endogenous explanatory variables computed from the simultaneous solution of the system at each iteration. Such iterations [illegible], as iterations of Brundy and Jorgenson (1971), [illegible]

information instrumental variables method, as shown in Hausman (1975). For linear models we have also

(40) $-\operatorname*{plim}_{T\to\infty} T^{-1}[\partial^2\mathcal{L}_T/\partial a\,\partial a']_{a_0} = \operatorname*{plim}_{T\to\infty} T^{-1}\tilde{G}'(\tilde{\Sigma}^{-1}\otimes I)\tilde{G}$

thus ensuring that matrix $\tilde{R}$ can consistently replace the Hessian in the case of hypotheses on linear models.

4. ALTERNATIVE ESTIMATORS OF FIML ASYMPTOTIC COVARIANCE MATRIX

Using the formulae of the previous section, we can build several estimators of FIML asymptotic covariance matrix. This can be done either for the model's unknown structural coefficients only, or for all unknown structural parameters, including the elements of the matrix Σ^{-1}. We may stack the estimated coefficients $\tilde{a}$ and the elements of the estimated $\tilde{\Sigma}^{-1}$ into a column vector of estimated parameters

(41) $\tilde{\delta} = \begin{bmatrix} \tilde{a} \\ \mathrm{vec}\tilde{\Sigma}^{-1} \end{bmatrix}$

Since $\tilde{\Sigma}^{-1}$ is symmetric, the vector $\tilde{\delta}$ would contain couples of elements equal to one another, corresponding to symmetric terms of $\tilde{\Sigma}^{-1}$ (of models with more than one stochastic equation). Any estimator of the whole information matrix would therefore be singular, as couples of rows (and columns) related to symmetric elements of $\tilde{\Sigma}^{-1}$ would be equal.

It is necessary to shorten properly the parameters vector $\tilde{\delta}$; we may still use the notation of equation (41), provided that $\mathrm{vec}\tilde{\Sigma}^{-1}$ is considered as the vector obtained by stacking only the columns of the upper triangular part of $\tilde{\Sigma}^{-1}$. In this way, if we consider the whole vector of parameters δ, its length is [illegible] structural coefficients, and m the number of stochastic equations; the whole information matrix (and the asymptotic covariance matrix) has dimensions $[n+m(m+1)/2]\times[n+m(m+1)/2]$. In those cases in which the asymptotic covariance matrix is related only to the structural coefficients, dimensions of the matrix will be $n\times n$.

We shall consider five different estimators of the asymptotic covariance matrix: since Ψ of equation (15) is related to the vector of coefficients only, we shall always aim at computing $\hat{\Psi}$ as an $n\times n$ matrix, but in two cases the computation will need to pass through the intermediate computation of the covariance matrix of the whole vector of structural parameters.

Generalized least squares type matrix

The inverse of matrix $\hat{R}$, as given in equation (38), can be used as an estimator of the asymptotic covariance matrix of the vector of estimated structural coefficients $\hat{a}$. Matrix $\hat{R}$ has the typical form of the matrix involved in generalized least squares estimation processes, being $\hat{G}$ the block diagonal matrix of observations, after cleaning explanatory endogenous variables of their component correlated with the error process (eq.34). It is by far the simplest to compute, among all traditional estimators. Its expression is also the most straightforward and natural estimator of the matrix R in Rothenberg and Leenders (1964, p.67), and its inverse is the most straightforward estimator of the lower bound for the asymptotic covariance matrix of consistent estimators of the coefficients, as it is given in Rothenberg (1973, p.67). Numerical exemplification of this estimator can be found in Hendry (1971), and in Hausman (1974).

Consistency of this estimator (under correct model's specification) is ensured only for linear models. The equality (40), in fact, does not hold

for nonlinear models, even if linear in the coefficients; the i,j-th block of the matrix $\hat{G}'(\hat{\Sigma}^{-1}\otimes I)\hat{G}$ should be replaced by a more complicated expression (see Amemiya, 1977, eqs. 3.14 and 4.10)

$$(42)\qquad \hat{G}_i'(\hat{\Sigma}^{-1}\otimes I)\hat{G}_j + [\sum_t(\partial g_{it}/\partial u_{jt})(\partial g_{jt}'/\partial u_{it})] - T^{-1}[\sum_t(\partial g_{it}/\partial u_{jt})][\sum_t(\partial g_{jt}'/\partial u_{it})].$$

Hessian of the concentrated log-likelihood

Equation (32) provides the i,j-th block of the Hessian matrix of the concentrated log-likelihood. As already observed, the first two terms on the right hand side are identically zero for linear models, and even models nonlinear in the variables. If computation is performed, as we do, at the values of coefficients a which maximize the likelihood, positive definiteness of this matrix is ensured.

Its inverse is one of the traditional estimators of the asymptotic covariance matrix of the vector of estimated structural coefficients of the model, $\hat{a}$, and consistency is ensured also for nonlinear models, under correct model's specification. Numerical applications of this estimator are rather frequent in the literature; see for example Chernoff and Divinsky (1953), and Klein (1969).

Equations (22-29), with the minus sign and summed over the sample period, provide the elements or the blocks of the $[n+m(m+1)/2]\times[n+m(m+1)/2]$ Hessian matrix of the unconcentrated log-likelihood

$$(43)\qquad -\partial^2 L_T/\partial p\partial p' = -\sum_t \partial^2 L_t/\partial p\partial p'$$

$$= -\begin{bmatrix} \partial^2 L_T/\partial a\partial a' & \partial^2 L_T/\partial a\partial(\mathrm{vec}\Sigma^{-1})' \\ \partial^2 L_T/\partial(\mathrm{vec}\Sigma^{-1})\partial a' & \partial^2 L_T/\partial(\mathrm{vec}\Sigma^{-1})\partial(\mathrm{vec}\Sigma^{-1})' \end{bmatrix}$$

The inverse of the matrix (43), computed at the maximum likelihood point, provides an estimator of the asymptotic covariance matrix of the whole vector of estimated structural parameters of the model. However, it must be recalled that, given the way in which the expression of Σ is substituted into the likelihood to obtain the concentrated likelihood, the first $n \times n$ block of the inverse of matrix (43) is equal to the inverse of the Hessian of the concentrated likelihood discussed above.

The inverse of the whole matrix (43) will be used together with the outer product of first derivatives to build a quasi maximum likelihood type covariance matrix.

Outer product of the unconcentrated first derivatives

Equations (20-22) provide the first derivatives of the unconcentrated log-likelihoods with respect to all the unknown structural form parameters. We may get an estimate of the whole information matrix by computing the outer product of the first derivatives

$$(44) \qquad \sum_t (\partial L_t/\partial p)(\partial L_t/\partial p')$$

$$= \sum_t \begin{bmatrix} (\partial L_t/\partial a)(\partial L_t/\partial a') & (\partial L_t/\partial a)[\partial L_t/\partial(\mathrm{vec}\Sigma^{-1})'] \\ [\partial L_t/\partial(\mathrm{vec}\Sigma^{-1})](\partial L_t/\partial a') & [\partial L_t/\partial(\mathrm{vec}\Sigma^{-1})][\partial L_t/\partial(\mathrm{vec}\Sigma^{-1})'] \end{bmatrix}$$

where all derivatives are computed at the value of a and Σ that maximize the likelihood. We then invert the whole $[n+m(m+1)/2]\times[n+m(m+1)/2]$ matrix, obtaining an estimate of the covariance matrix of the whole vector of estimated structural parameters, including coefficients and elements of $\hat{\Sigma}^{-1}$. The first $n \times n$ block of the inverse is an estimator of the asymptotic covariance matrix of the structural coefficients $\hat{a}$. Consistency is ensured for linear and non-linear models under correct model's specification.

We must notice that, even if we are interested in computing only the

$n \times n$ covariance matrix of the coefficients, here we build and invert the whole $[n+m(m+1)/2]\times[n+m(m+1)/2]$ information matrix, and then use only the first block of the inverse.

As a necessary condition for the invertibility of the whole matrix, the sample period length T should not be less than the total number of structural parameters $n+m(m+1)/2$ (see, for example, Hatanaka, 1978, p.333). This condition is more restrictive than the conditions which ensure existence and positive definiteness of the Hessian or of matrix $\hat{R}$, and become more and more restrictive with the enlargement of model's dimensions, since the minimum length of the sample period increases quadratically with the number of stochastic equations. This may be one of the reasons which prevent from a large use of this estimator in practice. Numerical exemplifications are, in fact, quite rare (see, for example, Artus et al., 1982 ,p.21).

Outer product of the concentrated first derivatives

The gradient of the concentrated log-likelihood, whose i-th subvector $\partial \ell_T/\partial a_i$ is given in equation (15), can be regarded as the sum of the T terms

$$(45) \qquad \partial g_{it}/\partial a_{i} - T(g_{it}\hat{f}_t')(\sum_t \hat{f}_t\hat{f}_t')_i^{-1}$$

each of which is equal to the corresponding derivative of the unconcentrated log-likelihood (20) when both are computed at the maximum likelihood point. We can, therefore, indifferently use (45) or (20) to compute the $n \times n$ outer product matrix as

$$(46) \qquad T^{-1} \sum_t (\partial L_t/\partial a)(\partial L_t/\partial a')$$

Berndt et al. (1974) propose the use of this matrix in gradient algorithms to maximize the likelihood function.

We can also use the inverse of this matrix as an estimator of the coefficients covariance matrix. The matrix is equal to the first block of matrix (44), but inversion would provide numerical results generally different from those obtained from the inversion of matrix (44), since the (1,2) and (2,1) blocks of matrix (44) are generally different from zero. Hatanaka (1976, pp.332 and 345) shows that, in the simultaneous equations case, such a difference does not vanish asymptotically, so that this estimator of the coefficients covariance matrix is in general inconsistent.

Quasi maximum likelihood type matrix

In quasi maximum likelihood estimation theory, the joint use of Hessian and outer product is recommended as a strongly consistent estimator of the asymptotic covariance matrix of model's parameters, even in cases of misspecification (see White, 1982, or Courieroux et al., 1984). If the model is correctly specified, this matrix would be asymptotically equivalent to the first three matrices discussed in this section. The matrix is obtained as

$$(47) \qquad (-\partial^2 L_T/\partial p \partial p')^{-1} [\sum_t (\partial L_t/\partial p)(\partial L_t/\partial p')] (-\partial^2 L_T/\partial p \partial p')^{-1}$$

and is an estimator of the $[n+m(m+1)/2]\times[n+m(m+1)/2]$ covariance matrix of all the model's parameters. Since we are interested in estimating the covariance matrix only of the structural coefficients, we compute only the first $n\times n$ block of matrix (47).

7. EXPERIMENTS ON SOME REAL-WORLD AND ARTIFICIAL MODELS

Experiments have been performed on five small linear models and on a medium nonlinear size model.

When the number of structural parameters is not greater than the sample period length, $n+m(m+1)/2\leq T$, the results are related to the estimates obtained from the historical sample. In this case each table presents on the leftmost column the forecasted values of the endogenous variables obtained with deterministic simulation of the model one-period ahead (static).

The other five columns display the standard errors of forecasts, due to errors in estimated coefficients only, obtained from applying equation (15) with the five different estimators of the covariance matrix of coefficients, $\hat{\Psi}/T$, discussed in section 6:

(1) generalized least squares type matrix;

(2) Hessian of the concentrated log-likelihood;

(3) outer product of the unconcentrated first derivatives;

(4) outer product of the concentrated first derivatives;

(5) quasi maximum likelihood type matrix.

When the number of structural parameters is too large, $n+m(m+1)/2>T$, so that the outer product matrix (44) is singular, the results are obtained with Monte Carlo. For each model, we start from a given set of parameters (coefficients and covariance matrix of the structural disturbances), we fix a sample period length and generate random values of the exogenous variables over the sample period. We use a multivariate normal generator, with given means and covariance matrix (taken from the historical sample). Independently of the exogenous, we then generate random values of the disturbance terms over the sample period, again with multivariate normal distribution, zero mean and the given covariance matrix. Values of the endogenous variables are finally computed with stochastic simulation over the sample period.

We now perform FIML estimation and compute, upon convergence at the values $\hat{a}$ and $\hat{\Sigma}^{-1}$ (the vector $\hat{p}$) which maximize the likelihood, all the five covariance matrices discussed in section 6.

We then generate random values of all the predetermined variables in the forecast period, and use the model with estimated coefficients, $\hat{a}$, to produce the forecast one-period ahead, and the five coefficients covariance matrices to compute the standard errors of forecasts. Of course, in the Monte Carlo case, forecasted values of the endogenous variables are not of particular interest and are, therefore, not displayed.

Table 1

Klein-I model

$$\begin{aligned}
C &= a_1 + a_2 P + a_3 P_{t-1} + a_4 (W1+W2) + u_1 \\
I &= a_5 + a_6 P + a_7 P_{t-1} + a_8 K_{t-1} + u_2 \\
W1 &= a_9 + a_{10}(Y+T-W2) + a_{11}(Y+T-W2)_{t-1} + a_{12} t + u_3 \\
Y &= C + I + G - T \\
P &= Y - W1 - W2 \\
K &= K_{t-1} + I
\end{aligned}$$

Number of equations = 6.
Number of stochastic equations $m = 3$.
Number of structural unknown coefficients $n = 12$.
Number of structural unknown parameters $n+m(m+1)/2 = 18$.

The meaning of the variables and empirical data for the U.S. economy 1921-1941 can be found, for example, in Rothenberg (1973, ch.5), while data for the predetermined variables in the forecast period have been taken from Goldberger et al. (1961). The sample estimation period is 1921-1941. One-period forecasts and standard errors due to errors in estimated coefficients are related to the year 1948.

Forecasts			Standard errors				
			(1)	(2)	(3)	(4)	(5)
C	=	76.5	1.37	1.89	5.43	2.70	2.33
I	=	6.27	1.01	1.86	3.99	1.82	3.14
W1	=	53.2	1.44	1.84	4.43	2.17	2.47
Y	=	90.9	2.25	3.62	9.28	4.42	5.29
P	=	26.1	1.19	2.13	5.19	2.37	3.45
K	=	204.	1.01	1.86	3.99	1.82	3.14

Table 2

Multiplier-accelerator model

$$C = a_1 + a_2 Y + a_3 C_{t-1} + u_1$$

$$E = a_4 + a_5 (Y - Y_{t-1}) + a_6 E_{t-1} + u_2$$

$$Y = C + E + G$$

Number of equations = 3.
Number of stochastic equations m = 2.
Number of structural unknown coefficients n = 6.
Number of structural unknown parameters n+m(m+1)/2 = 9.

The meaning of the variables and an example with empirical data for the U.S. economy (1949-1967) can be found in Dhrymes (1970, pp.533-534). The sample estimation period is 1949-1966. One-period forecasts and standard errors due to errors in estimated coefficients are related to the year 1967.

Forecasts	Standard errors (1)	(2)	(3)	(4)	(5)
C = 437.	3.53	3.59	5.57	3.85	3.73
E = 109.	5.50	5.10	12.1	9.61	5.18
Y = 635.	8.37	8.53	16.4	12.8	8.48

Table 3

A model for consumption and price of food

$$Q = a_1 P + a_2 + a_3 D + u_1$$

$$Q = a_4 P + a_5 + a_6 F + a_7 A + u_2$$

Number of equations = 2.
Number of stochastic equations m = 2.
Number of structural unknown coefficients n = 7.
Number of structural unknown parameters n+m(m+1)/2 = 10.

Model, variable names and a set of historical data for the U.S. economy 1922-1941 can be found in Kmenta (1971, pp.563-565). The two equations form a subset of the Girshick-Haavelmo's (1953) model. The sample estimation period is 1922-1940. One-period forecasts and standard errors due to errors in estimated coefficients are related to the year 1941.

Forecasts	Standard errors (1)	(2)	(3)	(4)	(5)
Q = 107	1.94	1.95	3.10	1.77	1.87
P = 114.	1.41	1.45	2.14	1.20	1.47

Table 4

A simple macro-economic model of the Italian economy

$$C = a_1 + a_2 Y + a_3 C_{t-1} + u_1$$

$$I = a_4 + a_5(Y - Y_{t-1}) + a_6 I_{t-1} + u_2$$

$$M = a_7 + a_8 I + a_9(Y - I) + u_3$$

$$Y = C + I + Z - M$$

Number of equations = 4.
Number of stochastic equations m = 3.
Number of structural unknown coefficients n = 9.
Number of structural unknown parameters $n+m(m+1)/2 = 15$.

The model, specifically designed for the purpose of analyzing the effects of current revisions in Italian national account series, is described in Trivellato and Rettore (1984). The estimation period is 1961-1979; ex-post forecasts and standard errors due to errors in estimated coefficients are related to the year 1980.

Forecasts			Standard errors				
			(1)	(2)	(3)	(4)	(5)
C	=	54229.	343.	421.	1790.	886.	328.
I	=	13913.	254.	307.	891.	530.	295.
M	=	17049.	257.	263.	375.	224.	325.
Y	=	85444.	467.	589.	2500.	1330.	456.

Table 5

Sitzia-Tivegna's model for the Italian economy

$$CPN = a_1 + a_2(WIT+WG+X2) + a_3(PIT+PAF) + a_4(PIT_{t-1}+PAF_{t-1}) + u_1$$

$$ILIT = a_5 + a_6 PIT_{t-1} + a_7 KOCC + a_8 ILIT_{t-1} + u_2$$

$$M = a_9 + a_{10}(CPN+ILIT) + a_{11} t + u_3$$

$$WIT = a_{12} + a_{13}(WIT+PIT) + a_{14} KOCC + a_{15} DUS70 + a_{16} WIT_{t-1} + u_4$$

$$KOCC = a_{17} + a_{18}(ILIT+ILIT_{t-1}+ILIT_{t-2}) + a_{19}(ILIT_{t-1}+2\,ILIT_{t-2}) + u_5$$

$$PIT = RNLCF - WIT - WG - PAF - X2$$

$$RNLCF = CPN + ILIT - M + WG - TI + X1$$

Number of equations = 7.
Number of stochastic equations m = 5.
Number of structural unknown coefficients n = 19.
Number of structural unknown parameters $n+m(m+1)/2 = 34$.

Model, meaning of the variables and data for the Italian economy 1952-1971 can be found in Sitzia and Tivegna (1975). Data over a sample period of 50 years have been generated with Monte Carlo.

Standard errors of forecasts (percentage)

		(1)	(2)	(3)	(4)	(5)
CPN	=	.466	.474	.997	.684	.469
ILIT	=	.931	.948	2.54	1.24	.912
SM	=	946	.964	2.29	1.17	1.00
WIT	=	.435	442	1.05	674	.372
KOCC	=	.533	.540	1.10	.557	.612
PIT	=	.607	.614	.907	586	787
RNLF	=	.252	.256	.540	.370	.266

Table 6

Klein-Goldberger model

Number of equations = 20.
Number of stochastic equations m = 16.
Number of structural unknown coefficients n = 54.
Number of structural unknown parameters $n+m(m+1)/2$ = 190.

For brevity's sake, the equations of the model are not reproduced. Model and data for the U.S. economy are described in Klein (1969). Data over a sample period with 300 observations have been generated with Monte Carlo.

Standard errors of forecasts (percentage)

		(1)	(2)	(3)	(4)	(5)
Cd	=	.465	.467	.762	.525	.456
Cn	=	.125	.125	.220	.136	.123
R	=	.528	.531	1.13	.716	.451
H	=	.312	.313	.555	.356	.295
Im	=	.373	.380	.663	.408	.391
X	=	.164	.165	.278	.179	.163
h	=	.193	.193	.366	.223	.189
W	=	.148	.149	.251	.159	.151
w	=	.075	.075	.134	.081	.077
p	=	.727	.740	1.17	.723	.779
I	=	.530	.541	.313	.533	.577
D	=	.255	.257	.485	.308	.245
Rs	=	.760	.771	1.29	.845	.777
Pc	=	1.66	1.67	2.95	1.31	1.63
Nw	=	.171	.171	.306	.197	.170
Y	=	.160	.161	.271	.170	.162
S	=	.250	.261	.535	.336	.271
Sc	=	16.3	16.3	29.9	19.5	16.5
[illegible]	=	.890	.895	1.42	.924	.782
Tr	=	.266	.267	.461	.300	.290

Table 7

Simple Keynesian model

$C = a_1 Y + u_1$

$Y = C + I$

Number of equations = 2.
Number of stochastic equations m = 1.
Number of structural unknown coefficients n = 1.
Number of structural unknown parameters $n+m(m+1)/2$ = 2.

FIML estimates are obtained for this model by means of simple algorithms, like indirect least squares, and simulation is fast enough even for very long sample periods. Data over a sample period of 30 years have been generated with Monte Carlo.

Standard errors of forecasts (percentage)

		(1)	(2)	(3)	(4)	(5)
C	=	16.6	16.6	19.1	17.6	17.1
Y	=	7.53	7.53	8.66	7.98	7.75

REFERENCES

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

Artus,P., G.Laroque, and G.Michel (1982), "Estimation of a Quarterly Model with Quantity Rationing". Paris: INSEE, discussion paper No.8209, presented at the European Meeting of the Econometric Society, Dublin.

Berndt,E.K., B.H.Hall, R.E.Hall, and J.A.Hausman (1974), "Estimation and Inference in Nonlinear Structural Models", Annals of Economic and Social Measurement 3, 653-665.

Bianchi,C., and G.Calzolari (1980), "The One-Period Forecast Errors in Nonlinear Econometric Models", International Economic Review 21, 201-208.

Bianchi,C., and G.Calzolari (1982), "Evaluating Forecast Uncertainty Due to Errors in Estimated Coefficients: Empirical Comparison of Alternative Methods", in Evaluating the Reliability of Macro-Economic Models, ed. by G.C.Chow, and P.Corsi. New York: John Wiley, 251-277.

Brundy,J.M., and D.W.Jorgenson (1971), "Efficient Estimation of Simultaneous Equations by Instrumental Variables", The Review of Economics and Statistics 53, 207-224.

Calzolari,G. (1981), "A Note on the Variance of Ex-Post Forecasts in Econometric Models", Econometrica 49, 1593-1595.

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

Calzolari,G., and F.P.Sterbenz (1983), "Efficient Computation of Reduced Form Variances in Nonlinear Econometric Models", presented at the European Meeting of the Econometric Society, Pisa.

Chernoff,H., and N.Divinsky (1953), "The Computation of Maximum-Likelihood Estimates of Linear Structural Equations", in Studies in Econometric Method, ed. by W.C.Hood and T.C.Koopmans. New York: John Wiley & Sons, Cowles Commission Monograph No.14, 236-302.

Dagenais,M.G. (1978), "The Computation of FIML Estimates as Iterative Generalized Least Squares Estimates in Linear and Nonlinear Simultaneous Equations Models", Econometrica 46, 1351-1362.

Dhrymes,P.J. (1970), Econometrics: Statistical Foundations and Applications. New York: Harper & Row.

Eisenpress,H., and J.Greenstadt (1966), "The Estimation of Nonlinear Econometric Systems", Econometrica 34, 851-861.

Fair,R.C. (1980), "Estimating the Expected Predictive Accuracy of Econometric Models", International Economic Review 21, 355-378.

Girshick,M.A., and T.Haavelmo (1953), "Statistical Analysis of the Demand for Food: Examples of Simultaneous Estimation of Structural Equations", in Studies in Econometric Method, ed. by W.C.Hood and T.C.Koopmans. New York: John Wiley & Sons, Cowles Commission Monograph No.14, 92-111.

Goldberger,A.S., A.L.Nagar and H.S.Odeh (1961), "The Covariance Matrices of Reduced-Form Coefficients and of Forecasts for a Structural Econometric Model", Econometrica 29, 556-573.

Gourieroux,C., A.Monfort, and A.Trognon (1984), "Pseudo Maximum Likelihood Methods: Theory", Econometrica 52, 681-700.

Hatanaka,M. (1978), "On the Efficient Estimation Methods for the Macro-Economic Models Nonlinear in Variables", Journal of Econometrics 8, 323-356.

Hausman,J.A. (1974), "Full Information Instrumental Variables Estimation of Simultaneous Equations Systems", Annals of Economic and Social Measurement 3, 641-652.

Hendry,D.F. (1971), "Maximum Likelihood Estimation of Systems of Simultaneous Regression Equations with Errors Generated by a Vector Autoregressive Process", International Economic Review 12, 257-272.

Klein,L.R. (1969), "Estimation of Interdependent Systems in Macroeconometrics", Econometrica 37, 171-192.

Kmenta,J. (1971), Elements of Econometrics. New York: The Macmillan Company.

McCarthy,M.D. (1972), "Some Notes on the Generation of Pseudo-Structural Errors for Use in Stochastic Simulation Studies", in Econometric Models of Cyclical Behavior, ed. by B.G.Hickman. New York: NBER, Studies in Income and Wealth No.3, 185-191.

Nissen,D.H. (1968), "A Note on the Variance of a Matrix", Econometrica 36, 603-604.

Parke,W.R. (1982), "An Algorithm for FIML and 3SLS Estimation of Large Nonlinear Models", Econometrica 50, 81-95.

Rothenberg,T.J. (1973), Efficient Estimation with A Priori Information. New Haven: Yale University Press, Cowles Foundation Monograph No.23.

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

Schink,G.R. (1971), "Small Sample Estimates of the Variance Covariance Matrix of Forecast Error for Large Econometric Models: the Stochastic Simulation Technique". University of Pennsylvania: Ph.D. dissertation.

Sitzia,B., and I.Tivegna (1975), "Un Modello Aggregato dell'Economia Italiana 1952-1971", in Contributi alla Ricerca Economica No.4. Roma: Banca d'Italia, 193-22[illegible].

ivellato.U., and E.Rettore (1984). "Preliminary Data Errors and Their Impact on the Forecast Error of Simultaneous Equations Models". Universita' di Padova: Dipartimento di Scienze Statistiche, discussion paper presented at the Fourth International Symposium on Forecasting, London.

ite,H. (1982), "Maximum Likelihood Estimation of Misspecified Models", Econometrica 50, 1-25.