A Simulation Study on FIML Covariance Matrix

Calzolari, Giorgio and Panattoni, Lorenzo

IBM Scientific Center, Pisa, Italy.

3 September 1984

A SIMULATION STUDY ON FIML COVARIANCE MATRIX

by Giorgio CALZOLARI *and* Lorenzo PANATTONI

Centro Scientifico IBM
via S.Maria 67
56100 Pisa (Italy)
tel. 050-47383

In econometric models, estimates of the asymptotic covariance matrix of FIML coefficients are traditionally computed in several different ways: with a *generalized least squares type* matrix; using the Hessian of the concentrated log-likelihood; using the outer product of the first derivatives of the log-likelihoods; with some suitable joint use of Hessian and outer product. The different alternative estimators are asymptotically equivalent in case of correct model's specification, but may produce large differences in the numerical application to small samples. The behaviour of the different estimators of the covariance matrix in *standardizing* or *normalizing* FIML estimated coefficients in the small samples is investigated in this paper. Monte Carlo experiments are performed on several small-medium size models, and some systematic behaviours are evidenced.

Preliminary version, not to be cited without permission. Comments welcome.

The authors have benefitted from discussions with D.A.Belsley, E.M.Cleur, A.Holly, G.Laroque, P.Mazodier, T.J.Rothenberg, N.E.Savin, P.Schmidt, A.Trognon, and H.K.Van Dijk. Responsibility, however, remains entirely with the authors.

CONTENTS

1. INTRODUCTION

Great effort has been placed, and is currently being placed in the literature on the investigation of parameters' estimators for simultaneous equation systems. Analysis based on the joint use of analytical and simulation methods seems quite useful for discriminating between estimators with the same asymptotic properties, as is the case of 2SLS and LIML in the recent works by Anderson, Kunitomo and Sawa (1982), and Morimune (1983). In any case, 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 three stage least squares to his model may look quite different from those obtained by full information maximum likelihood.

Let us now suppose that the model's structure and the length of the available data allow for the application of some full information estimation method, such as FIML, and that no problem of multiple maxima occurs in the feasibility region, so that the model builder gets a univocal vector of estimated coefficients whichever the starting point and whatever the maximization algorithm he is using.

He will now submit his model to some more or less sophisticated specification tests. In most cases he will start by testing coefficients one by one trying to provide evidence against the zero value hypothesis, and he will probably iterate from the introduction of slight changes in the model's specification to the reestimation of parameters until he will find "*t ratios*" greater than 2 for all the relevant coefficients. What will he use to compute these "*t ratios*" or other test statistics? No doubt that he will use the estimated asymptotic covariance matrix supplied by the computer program available to him and already used to get the estimate of coefficients, in other words:

(1) 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);

(2) the inverse of the Hessian of the concentrated log-likelihood, if his computer program is based on some *Newton-like* algorithm, like that of Eisenpress and Greenstadt (1966);

(3) the inverse of the outer product of the first derivatives of the log-likelihoods, if the program takes advantage of Berndt, Hall, Hall and Hausman's (1974) suggestions.

Although perfectly aware that these estimators of the coefficients covariance matrix are equivalent only for large samples, he would probably expect that even for his small sample the three groups of results had to be sufficiently close to one another, specially because the three matrices would be in any case computed *at the same point*. The numerical exemplification on a well known case study (Klein's model-I) should convince that this may not be the case.

Table 1

A case study: Klein's model-I. FIML estimates, hystorical data 1921-1941.

	Estimated coeffic.	Asymptotic standard errors (1)	(2)	(3)
Equation 1	18.34	2.485	4.626	12.88
	-.2324	.3120	.5806	1.931
	.3857	.2174	.3017	1.083
	.8018	.0359	.0445	.0842
Equation 2	27.26	7.938	9.535	21.47
	-.8010	.4914	.8402	2.334
	1.052	.3525	.4244	1.404
	-.1481	.0299	.0468	.0992
Equation 3	5.794	1.804	3.241	4.645
	.2341	.0488	.0950	.0953
	.2847	.0452	.0629	.0617
	.2348	.0345	.0565	.0776

All the standard errors (but one) computed from the outer product (3)

are larger (double or even triple in several cases) than those computed from the Hessian (2), and the latter on their turn are all considerably larger than those computed with the *generalized least squares type* matrix (1). If we perform a test based on the rough computation of the *"t ratios"* for all coefficients, the hypothesis of zero coefficient could be rejected for most of them if the standard errors had been computed with the *generalized least squares type* matrix (1), but it could not be rejected for most of them if the standard errors had been computed from the outer product (3).

A result like this might be used to bring evidence against the hypothesis of correct model's specification, as in White's test (1982) (see also Hall, 1983). However, this is not necessarily so in the small sample case. Even in case of exact model's specification (and of exact specification of the disturbance process), as it is the case of an artificial model in a Monte Carlo experiment, it is quite frequent that we encounter a phenomenon like the one just observed: for the sample sizes occurring in practice (and even for considerably longer samples) the signs of the differences already encountered are practically almost systematic. In other words, standard errors computed from the outer product are almost systematically larger than those computed from the Hessian, and the latter are, on their turn, almost systematically larger than those computed from a *generalized least squares type* matrix.

The first purpose of this paper is, therefore, that of warning those model builders who are in the fortunate position of affording FIML estimation: the choice of the estimator of the coefficients asymptotic covariance matrix is not as neutral as it might be expected.

In principle each of the *"t ratio"* statistics should be investigated, so to produce tables of percentiles for the different cases; the critical values should not always be ±2, but it should be chosen case by case from the appropriate small sample distribution. This would introduce additional

burden into the already burdensome process of FIML estimation, and model builders would probably continue in their habit of using ±2 as the critical values good for all circumstances.

As a natural alternative, and this is the second purpose of the paper, one could try to investigate if one of the different estimators of the covariance matrix could in some way be preferred to the others. In particular a result quite useful for model builders would be to know which of the estimators is best for *standardizing* or *normalizing* the coefficient estimates, in such a way that one might reliably continue in using the standard normal as a reference distribution for his test statistics (as is for the "*t ratio*" in the linear regression model for the sample sizes occurring in practice).

2. MAIN ASSUMPTIONS AND NOTATIONS

Although experimentations will be performed only on linear or linearized models, in deriving the analytic expressions of the covariance matrices we follow the general notation of Amemiya (1977). It is particularly useful in case of linearized models, since no explicit linearization has to be preliminarily done, but it proves to be quite useful also for linear models, where the traditional notation, involving the use of large sparse matrices, increases the risk of computer programming errors. Reference to Amemiya's paper should also be done for detail on the underlying assumptions. Let the simultaneous equation model be represented as

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

where y_t is the $m\times 1$ 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 $m \times 1$ 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 $(n \times 1)$ vector of unknown structural coefficients of the system will be indicated as $a=(a_1',a_2',\ldots,a_m')'$.

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

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

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

$$(3) \qquad L_T = \sum_t L_t .$$

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

$$(4) \qquad \partial L_t/\partial a_i = \partial g_{it}/\partial u_{it} - g_{it} f_t' \sigma^i$$

$$(5) \qquad \partial L_t/\partial(\Sigma^{-1}) = 1/2\, \Sigma - 1/2\, f_t f_t'$$

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

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

Using $\partial g_{ijt}/\partial u_{it}=(\partial g_{ijt}/\partial y_t')(\partial f_t/\partial y_t')_i^{-1}$, further differentiation of (4) gives

(7) $\quad \partial^2 L_t/\partial a_i\partial a_i' = \partial g_{ijt}/\partial u_{it} - (\partial g_{it}/\partial u_{jt})(\partial g_{jt}'/\partial u_{it})$

$\qquad\qquad - g_{ijt}f_t'\sigma^i - \sigma^{ij}g_{it}g_{jt}'$

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

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

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

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

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

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

Under standard assumptions, by equating to zero the first order derivatives of the unconcentrated log-likelihood with respect to Σ^{-1} (5), and substituting back in (2) and (3), we get the concentrated log-likelihood function

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

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

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

which is equal to $\partial L_T/\partial a_j$, obtained from summing (4) over time, provided that covariance parameters are replaced with their FIML estimates.

Further differentiation of (15), with respect to the structural coefficients of equation j, gives the i,j-*th* block of the Hessian matrix of the concentrated log-likelihood

$$(16) \quad -\partial^2 \ell_T/\partial a_i \partial a_j' = -\sum_t \partial g_{ijt}/\partial u_{it} + T\Big(\sum_t g_{ijt} f_t'\Big)\Big(\sum_t f_t f_t'\Big)_i^{-1} + \Big[\sum_t (\partial g_{it}/\partial u_{jt})(\partial g_{jt}'/\partial u_{it})\Big] + T\Big(\sum_t f_t f_t'\Big)_{ij}^{-1}\Big(\sum_t g_{it} g_{jt}'\Big) - T\Big(\sum_t g_{it} f_t'\Big)\Big(\sum_t f_t f_t'\Big)_j^{-1}\Big(\sum_t f_t f_t'\Big)_{i'}^{-1}\Big(\sum_t f_t g_{jt}'\Big) - T\Big(\sum_t f_t f_t'\Big)_{ij}^{-1}\Big(\sum_t g_{it} f_t'\Big)\Big(\sum_t f_t f_t'\Big)^{-1}\Big(\sum_t f_t g_{jt}'\Big).$$

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 (7) and the first two terms on the right hand side of equation (16) 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 $\partial g_{it}/\partial u_{it} = (\partial g_{it}/\partial y_t')(\partial f_t/\partial y_t')_i^{-1}$, this differentiation could even be performed analytically without any particular difficulty. The use of equations (7-13) and (16) 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 instrumental variables approach.

Introducing the $T \times m$ 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 (15) can be rewritten as

$$(17) \qquad \partial \ell_T / \partial a_i = [T^{-1} \sum_t (\partial g_{it} / \partial u_t') F' - G_i'] \; F (T^{-1} F'F)_i^{-1} .$$

We define, now,

$$(18) \qquad \hat{G}_i = G_i - T^{-1} F \sum_t (\partial g_{it}' / \partial u_t)$$

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 a, we have

$$(19) \qquad T^{-1} \hat{F}' \hat{F} = \hat{\Sigma} .$$

To compute FIML estimates, we must solve the nonlinear system obtained by equating to zero the gradient of the concentrated likelihood (17). Equating (17) to zero, and combining all equations for $i=1,2,\ldots,m$, we get

$$(20) \qquad \hat{C}'(\hat{\Sigma}^{-1} \otimes I)\, vec\hat{F} = 0$$

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

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 $vecF$ as a function of the coefficients vector, a. The simple iterative method which results is

$$(21) \qquad \hat{a}^{k} = \hat{a}^{k-1} - [\hat{C}'(\hat{\Sigma}^{-1} \otimes I)C]^{-1}\hat{C}'(\hat{\Sigma}^{-1} \otimes I)vec\hat{F}.$$

A more convenient iterative method is obtained if the square matrix which appears in brackets on the right hand side of (21) is replaced by the matrix

$$(22) \qquad \hat{R} = [\hat{C}'(\hat{\Sigma}^{-1} \otimes I)\hat{C}]$$

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, $\partial g_{it}/\partial u_{jt}$ is no more time varying; if the model is

$$(23) \qquad Ay_t + Bx_t = u_t$$

then the vector $\partial g_{it}/\partial u_{jt}$, 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 $\hat{C}$ would simply be the values of the endogenous explanatory variables computed from the simultaneous solution of the system at each iteration. Each iteration is, therefore, an iteration of Brundy and

Jorgenson's (1971) full information instrumental variables method, as shown in Hausman (1974). For linear models we have also

$$(24) \qquad -\underset{T\to\infty}{plim}\, T^{-1}[\partial^2 \ell_T/\partial a \partial a']_{a_0} = \underset{T\to\infty}{plim}\, T^{-1}\hat{C}'(\hat{\Sigma}^{-1}\otimes I)\hat{C}$$

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

3. ESTIMATORS OF THE ASYMPTOTIC COVARIANCE MATRIX

Using the formulae of the previous section, we can build several estimators of the 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 $\hat{a}$ and the elements of the estimated $\hat{\Sigma}^{-1}$ into a column vector of estimated parameters

$$(25) \qquad \hat{p} = \begin{bmatrix} \hat{a} \\ vec\hat{\Sigma}^{-1} \end{bmatrix}.$$

Since $\hat{\Sigma}^{-1}$ is symmetric, the vector $\hat{p}$ would contain couples of elements equal to one another, corresponding to symmetric terms of $\hat{\Sigma}^{-1}$ for 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 $\hat{\Sigma}^{-1}$ would be equal.

It is necessary to shorten properly the parameters vector $\hat{p}$; we may still use the notation of equation (25), provided that $vec\hat{\Sigma}^{-1}$ is considered as the vector obtained by stacking only the columns of the upper triangular part of $\hat{\Sigma}^{-1}$. In this way, if we consider the whole vector of

parameters $\hat{p}$, its length is $n+m(m+1)/2$, n being the number of unknown 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 nine different estimators of the asymptotic covariance matrix: six will be related to the vector of coefficients only, while three will be related to the whole vector of structural parameters.

3.1. Generalized least squares type matrix

The inverse of matrix $\hat{R}$, as given in equation (22), can be used as an estimator of the asymptotic covariance matrix of the vector of estimated structural coefficients $\hat{a}$ (without the elements of matrix $\hat{\Sigma}^{-1}$). It has been used to compute the first column of standard errors in the example of Table 1. Matrix $\hat{R}$ has the typical form of the matrix involved in *generalized least squares* estimation processes, being $\hat{C}$ the block diagonal matrix of observations, after cleaning explanatory endogenous variables of their component correlated with the error process (eq.18). 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 (24), 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)

$$(26) \qquad \hat{G}_i'(\hat{\Sigma}^{-1}\otimes I)\hat{G}_j + [\sum_t (\partial g_{it}/\partial u_{jt})(\partial g_{jt}'/\partial u_{it})] - T^{-1}[\sum_t (\partial g_{it}/\partial u_{jt})][\sum_t (\partial g_{jt}'/\partial u_{it})]$$

Numerical examples related to nonlinear models, however, are not given in this paper.

3.2. Hessian of the concentrated log-likelihood

Equation (16) 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, but the latter will not be considered in this paper). 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, $\hat{a}$ (without the elements of matrix $\hat{\Sigma}^{-1}$), 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). This matrix has been used to compute the second column of standard errors in the example of Table 1.

3.3. Hessian of the unconcentrated log-likelihood

Equations (7-13), 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

$$(27) \qquad -\partial^2 L_T/\partial p\partial p' = -\sum_t \partial^2 L_t/\partial p\partial p'$$

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

The inverse of the matrix (27), 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. Consistency is ensured for linear and nonlinear models under correct model's specification. Numerical applications of this estimator are quite rare in the literature; an example is given in Calzolari (1983).

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 (27) is equal to the inverse of the Hessian of the concentrated likelihood discussed in section (3.2). Therefore there is no difference between the two estimators as far as the covariance matrix of coefficients is concerned, but only the inverse of (27) can be used to estimate the asymptotic variances-covariances related to the elements of Σ^{-1}.

3.4. Outer product of the unconcentrated first derivatives

Equations (4-6) provide the first derivatives of the unconcentrated log-likelihoods with respect to all the unknown structural form parameters. We may get an estimate the whole information matrix by computing the outer product of the first derivatives

$$(28) \qquad \sum_t (\partial L_t/\partial p)(\partial L_t/\partial p')$$

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

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 nonlinear 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, Laroque and Michel, 1982,p.21). This matrix has been used to compute the third column of standard errors in the example of Table 1.

3.5. 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

$$(29) \qquad \partial g_{it} / \partial u_{it} \cdot T(g_{it} f_t')\left(\sum_t f_t f_t'\right)_i^{-1}$$

each of which is equal to the corresponding derivative of the unconcentrated log-likelihood (4) when both are computed at the maximum likelihood point. We can, therefore, indifferently use (29) or (4) to compute the $n \times n$ outer product matrix as

$$(30) \qquad T^{-1} \sum_t (\partial L_t / \partial a)(\partial L_t / \partial a').$$

Berndt, Hall, Hall and Hausman (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 (28), but inversion would provide numerical results generally different from those obtained from the inversion of matrix (28), since the (1,2) and (2,1) blocks of matrix (28) are generally different from zero. Hatanaka (1978, 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. Our simulation results, however, show that in the small or medium sizes of the samples usually occurring in practice (and even for longer samples) the behaviour of this estimator is often better than the behaviour of some other estimators, theoretically more appealing.

3.6. Quasi maximum likelihood type matrices

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, and Gourieroux, Monfort and Trognon, 1982). If the model is correctly specified, this matrix would be asymptotically equivalent to all the other matrices discussed in section 3 (except 3.5, as already observed). The matrix is obtained as

$$(31) \qquad (-\partial^2 L_T/\partial p \partial p')^{-1} \{ \sum_t (\partial L_t/\partial p)(\partial L_t/\partial p') \} (-\partial^2 L_T/\partial p \partial p')^{-1}$$

and is an estimator af the $[n+m(m+1)/2]\times[N+m(m+1)/2]$ covariance matrix of all the model's parameters. If we are interested in estimating the covariance matrix only of the structural coefficients, we can compute only the first $n\times n$ block of matrix (31).

Alternatively, we might use in the first and third term of (31) the $n\times n$ Hessian of the concentrated likelihood, whose inverse is equal to the first block of the inverted Hessian of the unconcentrated likelihood. For the intermediate term we can adopt the following method, which surely avoids the inconsistency problems that might arise by simply using equation (30). We compute the outer product matrix of unconcentrated first derivatives, invert the whole matrix, take the first $n\times n$ block of the inverse, re-invert such an $n\times n$ matrix and use the result to replace the intermediate outer product in (31). Under correct model's specification, we get a different $n\times n$ consistent estimator of the coefficients covariance matrix.

4. DESIGN OF THE MONTE CARLO EXPERIMENTS

A wide set of Monte Carlo experiments has been performed on several small-medium size models. The models, taken from the literature, maintain the structure of real world models.

For each model, we start from a given set of *true* parameters (coefficients and covariance matrix of the structural disturbances, held fixed over all the replications), 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). The results, however, should not be particularly sensitive to the choice of the distribution used to generate exogenous variables, provided that their covariance matrix is finite, and in fact the results do not significantly differ from the ones obtained with fixed values of the exogenous in all the replications (like those discussed in Calzolari and Panattoni, 1983).

Independently of the exogenous, we then generate random values of the disturbance terms, u_t, over the sample period. Obviously this distribution *must* be multivariate normal with 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 a least squares estimation of the coefficients for all the equations of the model, in such a way as to get a *reasonably good* starting point for the maximization process. Then we perform the first iterations of FIML using a gradient algorithm based on matrix R, which usually proved to be computationally more efficient (see Calzolari and Panattoni, 1983). Intermediate iterations are then made using a *Newton like* algorithm based on the Hessian of the concentrated likelihood followed by a search of the maximum in the chosen direction. The last iterations are performed using Newton's method. Since the purpose of this paper is

not to compare the computational efficiency of different maximization algorithms, but to compare the behaviour of different estimators of the covariance matrix, all computed at the same point, it seemed worthwhile to ensure an extremely accurate computation of the maximum point by choosing a very tight convergence criterion: 10^{-9} as a relative tolerance on all coefficients.

Upon convergence we compute, at the values $\hat{a}$ and $\hat{\Sigma}^{-1}$ (the vector $\hat{p}$) which maximize the likelihood, all the covariance matrices discussed in section 3. Six of these matrices have dimensions $n \times n$ and can be used to standardize or normalize the vector of coefficients errors $\hat{a}-a$ (where a is the vector of *true* coefficients). The other three matrices have dimensions $[n+m(m+1)/2] \times [n+m(m+1)/2]$ and can be used to standardize or normalize the vector of all structural parameters errors $\hat{p}-p$ (where p includes a and the elements of Σ^{-1}). This is the last step of each Monte Carlo replication.

For each model and for each sample period length we perform a few hundred replications of the Monte Carlo process (500 for short and medium lengths of the sample periods, 200 for the longest cases), obtaining six small sample distributions for each of the standardized coefficients, and three small sample distributions for each of the standardized parameters. For each coefficient or parameter we may now investigate how closely its standardized distributions approximates the standard normal, or how fast it converges to the standard normal, when the sample period length increases.

The number of distributions which should be examined becomes quickly very large. For a small model like Klein-I, in fact, we have 6 distributions for each of the 12 structural coefficients, and 3 more distributions for each of the 18 structural parameters. For a medium size model, like Klein-Goldberger, there would be 6 distributions for each of the 54 coefficients, and 3 more distributions for each of the 190

structural parameters. All these distributions must be re-examined for several lengths of the sample period (and possibly for different choices of the *true* structural parameters and of the distribution of the exogenous variables). This would make the analysis extremely dispersive, or even impossible in practice. We can make it feasible in two different ways.

We may select one or two parameters of interest and confine the analysis to them, disregarding all the others. Alternatively, we can combine all the coefficients errors, or parameters errors, into a single random variable, and examine how close the 9 small sample distributions are to their asymptotic distribution. We have followed this second approach.

Let $\hat{\Psi}$ be one of the $n \times n$ estimators of the coefficients covariance matrix; then

$$(32) \qquad (\hat{a}-a)' \, \hat{\Psi}^{-1} \, (\hat{a}-a)$$

is asymptotically distributed as χ^2 with n degrees of freedom. Analogously, if $\hat{\Phi}$ is one of the $[n+m(m+1)/2] \times [n+m(m+1)/2]$ estimators of the parameters covariance matrix, then

$$(33) \qquad (\hat{p}-p)' \, \hat{\Phi}^{-1} \, (\hat{p}-p)$$

is asymptotically distributed as χ^2 with $n+m(m+1)/2$ degrees of freedom.

For each model and each sample period length, we get 6 distributions to be compared with $\chi^2_{(n)}$, and 3 distributions to be compared with $\chi^2_{[n+m(m+1)/2]}$. In the tables which follow, the six variables whose distribution is asymptotically approximated by $\chi^2_{(n)}$ will be indicated by numbers 1-6 as follows.

1 - In equation (32) the covariance matrix has been computed from the *generalyzed least squares type* matrix $\hat{R}$, as in section 3.1.

2 - The covariance matrix has been computed from the Hessian of the

concentrated log-likelihood, as in section 3.2.

3 - The covariance matrix is the first $n \times n$ block of the inverse of the outer product of the unconcentrated first derivatives, as in section 3.4.

4 - The covariance matrix has been computed from the outer products of the concentrated first derivatives, as in section 3.5.

5 - The first $n \times n$ block of the $[n+m(m+1)/2] \times [n+m(m+1)/2]$ *quasi maximum likelihood type* matrix (31) is used.

6 - The last version of *quasi maximum likelihood type* matrices described in section 3.6 is used.

The three variables whose distributions are asymptotically approximated by $\chi^2_{[n+m(m+1)/2]}$ will be indicated by numbers 7, 8 and 9, as follows.

7 - The covariance matrix to be used in equation (33) has been computed from the Hessian of the unconcentrated likelihood, as in section 3.3.

8 - The outer product of unconcentrated first derivatives has been used as in equation (28).

9 - The whole $[n+m(m+1)/2] \times [n+m(m+1)/2]$ *quasi maximum likelihood type* matrix (31) has been used for computation of the covariance matrix to be inserted into (33).

5. EXPERIMENTS ON SMALL-MEDIUM SIZE MODELS

Monte Carlo experiments have been performed on five small linear models and on a linearized medium size model.

Random data have been generated for several sample period lengths. For each small linear model, the figures display results related to a short, a medium length and a relatively long sample period. Simulations over longer sample periods are not reproduced, since most of the

cumulative distributions collapse over the corresponding χ^2 distribution. However, simulations could not be done for sample periods shorter than the total number of parameters given the singularity problems that would arise for the outer product matrices, as observed in section 3.4. This fact, and the computational problems arising from the joint use of medium size models and long sample periods, limited the experiments with the linearized medium size model to a couple of cases in which the sample period was not many times longer than the total number of parameters.

All these results are summarized in the figures 1 to 17. Each set of figures relevant to a given model is preceded by a table containing the principal characteristics of the model, the bibliographic reference and the model specification (this last has been omitted for the medium size model).

6. REMARKS ON SOME SYSTEMATIC BEHAVIOURS

Consistency of the estimators for the asymptotic covariance matrices ensures that, increasing the sample period lengths, the cumulative distributions in Figures 1-17 collapse over the corresponding χ^2 curves. This does not hold for distribution 4 (outer product of concentrated first derivatives), as it is particularly evident from Figures 3, 6 and 17. However, for sample periods not very much longer than the total number of parameters, the outer product of concentrated derivatives (distribution 4) behaves better than the *quasi maximum likelihood type* matrices (distributions 5,6 and 9; see, for example, Figures 4, 11, 13 and 14). Of course this does not touch the merits of these last matrices in the very important case of misspecification.

For the smaller models, the outer product of unconcentrated derivatives (distribution 3) seems to behave better than all the other matrices. This is no more so for the larger models, but still seems to

hold when all the structural parameters are considered (distribution 8). However, since model builders are usually interested in tests related to coefficients and rarely in the covariance parameters, our experiments do not provide a univocal indication of superiority of one estimator over all the others. This problem still needs further investigation.

Some other behaviours, which occur in all the experiments, might be probably considered systematic.

First of all we notice that the random variables obtained from the *generalized least squares type* matrices and from the Hessians have distributions (1 and 2) rather close to one another, but when the two curves are distinguishable, the latter curve is systematically on the left. This means that matrix $\hat{R}$ tends to be larger (in matrix sense) than the Hessian and therefore, after inversion, that the variances of coefficients computed from matrix $\hat{R}$ tend to be smaller than those computed from the Hessian. The fact that the two curves are rather close to each other suggests that the standard errors of coefficients, even if different, are *on the average* not so different as they are for Klein-I model with historical data (Table 1).

Both curves are systematically on the right of the χ^2 distribution, suggesting that both matrices provide estimates of the covariance matrix which tend to be too small.

If we now extend the comparison to the outer product of the unconcentrated first derivatives, we notice that the corresponding distribution is further left shifted from the Hessian, both when only the coefficients are considered (3), and when all the structural parameters are considered (8). This suggests that the covariance matrix computed from the outer product tends to be larger than that computed from the Hessian. The former, when only coefficients are considered, gives a curve (3) which is neither systematically on the left, nor systematically on the right of the theoretical χ^2 for all sample sizes, except for the

largest models, for which it is systematically on the left (thus suggesting that variances computed from the outer products tend to be too large).

The behaviour of the *quasi maximum likelihood* type matrices is a natural consequence of the behaviour of Hessian and outer product matrices. Roughly speaking, if we look at equation (31) and remind that the inverted Hessian tends to be smaller than the inverted outer product, and that the outer product (not inverted) tends to be smaller than the Hessian (not inverted), as a result of the product (31) we get a matrix which tends to be even smaller than the inverted Hessian. The corresponding curves (5, 6 and 9) are, therefore, expected to be right shifted with respect to those obtained from the Hessian. The fact that they are so much shifted and far from the theoretical χ^2 is, however, remarkable and would be worthwhile further investigation.

6.1. Hessian versus generalized least squares type matrix

The behaviour of the Hessian matrix when compared with the *generalized least squares type* matrix $\hat{R}$ can be partially explained as follows.

In case of the linear system (23), let

$$
\begin{aligned}
&\Pi = -A^{-1}B && \text{and } \hat{\Pi} = -\hat{A}^{-1}\hat{B} \text{ its FIML estimate} \\
&\hat{M} = T^{-1}\sum_t x_t x_t' && \text{and } M \text{ its probability limit} \\
&\hat{N} = T^{-1}\sum_t x_t u_t' && \text{with zero probability limit} \\
&\hat{\Sigma} = T^{-1}\sum_t u_t u_t' && \text{and } \Sigma \text{ its probability limit.}
\end{aligned}
\qquad (34)
$$

The matrix G_i, which consists of T rows and as many columns as the number of explanatory variables in the *i-th* equation, can be obtained by properly selecting columns of the matrix

$$(35)\qquad \begin{bmatrix} y_1' ; & x_1' \\ y_2' ; & x_2' \\ \vdots & \vdots \\ y_T' ; & x_T' \end{bmatrix} = \begin{bmatrix} x_1' \\ x_2' \\ \vdots \\ x_T' \end{bmatrix} [\Pi' \; ; \; I] + \begin{bmatrix} u_1' \\ u_2' \\ \vdots \\ u_T' \end{bmatrix} [A'^{-1} \; ; \; 0]$$

and the matrix $\hat{C}_i$, of equation (18), can be obtained from properly selecting columns of the matrix

$$(36)\qquad \begin{bmatrix} x_1' \\ x_2' \\ \vdots \\ x_T' \end{bmatrix} [\hat{\Pi}' \; ; \; I]$$

The i,j-*th* block of the matrix $\hat{R}$ can be obtained from properly selecting rows (corresponding to the explanatory variables of equation i) and columns (corresponding to the explanatory variables of equation j) of the matrix

$$(37)\qquad T \begin{bmatrix} \hat{\Pi}\hat{M}\hat{\Pi}' & \hat{\Pi}\hat{M} \\ \hat{M}\hat{\Pi}' & \hat{M} \end{bmatrix} \hat{a}^{ij}.$$

The formulas which follow become simpler if we avoid to represent matrices like (37) in partitioned form. This can be simply accomplished by properly augmenting the vector of endogenous variables of the system with the inclusion of all variables which multiply structural coefficients in any stochastic equation and that are either predetermined or functions of endogenous variables. Such a representation of model (23) simply needs the addition of definitional equations, and all the unknown structural coefficients become elements of the matrix A, while no unknown coefficient appears any more in the matrix B. Since all explanatory variables on the right hand side of any stochastic equation are, now, formally represented as endogenous, we can drop the rightmost part of the partitioned matrices (35) and (36), while the whole matrix (37) can be represented as

$$(38) \qquad T\,[\hat{\Pi}\hat{M}\hat{\Pi}']\hat{\omega}^{ij}.$$

As above, we must select rows and columns of matrix (38) in order to build the *i,j-th* block of the matrix $\hat{R}$. The *i,j-th* block of the Hessian of the concentrated log-likelihood (16), $-\partial^2\ell_T/\partial a_i\partial a_j'$, calculated at the FIML estimate of a, can be built starting from the last four terms on the right hand side of equation (16) (the first two terms are zero in our case). As above, for any i and j, we must select the same rows and columns of the matrices which will be given below.

- *3rd* term of the concentrated Hessian:

$$(39) \qquad T\,[\hat{A}^{-1}i_j i_j'\hat{A}'^{-1}]$$

where i_i and i_j are the *i-th* and *j-th* columns of the $m\times m$ unit matrix.

- *4th* term of the concentrated Hessian:

$$(40) \qquad T\,[\hat{\Pi}\hat{M}\hat{\Pi}' + \hat{\Pi}\hat{N}\hat{A}'^{-1} + \hat{A}^{-1}\hat{N}'\hat{\Pi}' + \hat{A}^{-1}\hat{\Sigma}\hat{A}'^{-1}]\,\hat{\omega}^{ij}.$$

- *5th* term of the concentrated Hessian:

$$(41) \qquad -T\,[\hat{\Pi}\hat{N}\hat{\omega}^j\hat{\omega}^{i\prime}\hat{N}'\hat{\Pi}' + \hat{\Pi}\hat{N}\hat{\omega}^j i_i'\hat{A}'^{-1} + \hat{A}^{-1}i_i\hat{\omega}^{i\prime}\hat{N}'\hat{\Pi}' + \hat{A}^{-1}i_i i_j'\hat{A}'^{-1}]$$

where use has been done of $\hat{\Sigma}\hat{\omega}^j=i_j$ and $\hat{\omega}^{i\prime}\hat{\Sigma}=i_i'$.

- *6th* term of the concentrated Hessian:

$$(42) \qquad -T\,[\hat{\Pi}\hat{N}\hat{\Sigma}^{-1}\hat{N}'\hat{\Pi}' + \hat{\Pi}\hat{N}\hat{A}'^{-1} + \hat{A}^{-1}\hat{N}'\hat{\Pi}' + \hat{A}^{-1}\hat{\Sigma}\hat{A}'^{-1}]\,\hat{\omega}^{ij}.$$

When summing the matrices (39), (40), (41) and (42), several terms cancel; we get

$$(43) \qquad T\hat{\Pi}\hat{M}\hat{\Pi}'\hat{\sigma}^{ij} - T\hat{\Pi}\hat{N}\hat{\Sigma}^{-1}\hat{N}'\hat{\Pi}'\hat{\sigma}^{ij} - T[\hat{\Pi}\hat{N}\hat{\sigma}^{j}\hat{\sigma}^{i\prime}\hat{N}'\hat{\Pi}'+\hat{\Pi}\hat{N}\hat{\sigma}^{j}i_{j}'\hat{A}'^{-1}+\hat{A}^{-1}i_{j}\hat{\sigma}^{i\prime}\hat{N}'\hat{\Pi}'].$$

The first term of (43) is the matrix (38). The other terms, given the presence in each of them of the matrix $\hat{N}$, whose probability limit is zero, asymptotically vanish as expected. For small samples, however, the second term of (43) contributes to make the resulting matrix smaller (in the usual matrix sense); in fact, if we build the entire matrix whose *i,j-th* block is obtained from selecting rows and columns of the second term of (43), we would get, by defining an appropriate matrix $\hat{Z}$, a matrix of the form $\hat{Z}'(\hat{\Sigma}^{-1}\otimes I)\hat{Z}$, that is positive semidefinite, and this matrix should be subtracted from $\hat{R}$. The last term of (43), in brackets, has, however, a quite different behavior. Each block has, in fact, a maximum rank equal to one, since the block is obtained as the product of a column vector with a row vector. Moreover, several elements of the matrix, made up of these blocks, would be zero if calculated upon convergence of the FIML estimation and in particular all the elements of the diagonal blocks would be zero. This follows from considering that, from equation (20), the subvector of the gradient, corresponding to the coefficients of the *i-th* equation, could be obtained from properly selecting elements from the vector $\hat{\Pi}\hat{N}\hat{\sigma}^{i}$; since this subvector is zero at the optimum, the matrix in brackets on the right hand side of equation (43) is zero when $i=j$.

We must now invert $\hat{R}$ and $-\partial^2\ell_T/\partial a_i\partial a_j'$ to get two of the estimators of the asymptotic covariance matrix of structural coefficients. If the Hessian were derived only from the first two terms of (43), the inverted Hessian would be always greater than $\hat{R}^{-1}$. Although this inequality is not exact, given the presence of the last term in (43), it holds in most cases.

6.2. Outer product versus Hessian

The systematic behaviour of the Hessian matrix versus the outer product of unconcentrated first derivatives has a simple explanation in a particular single equation case. Let y_t be distributed as $N(0,\sigma^2)$. We can represent the model with the notation (1) as

(44) $$y_t - a = u_t$$

and the vector of unknown parameters as

(45) $$p = [a \ , \ \sigma^{-2}]'.$$

In this model $g_{it}=-1$ for all t, and the sum of residuals is zero at the maximum likelihood point. Therefore, from equations (27) and (6-13), it follows that the Hessian of the unconcentrated log-likelihood is simply

(46) $$-\partial^2 L_T/\partial p \partial p' = \begin{bmatrix} T\hat{\sigma}^{-2} & 0 \\ 0 & T\hat{\sigma}^4/2 \end{bmatrix}$$

while the outer product of first derivatives of the unconcentrated log-likelihoods becomes

(47) $$\sum_t (\partial L_t/\partial p)(\partial L_t/\partial p') = \begin{bmatrix} T\hat{\sigma}^{-2} & \sum_t \hat{u}_t^3/2 \\ \sum_t \hat{u}_t^3/2 & (\sum_t \hat{u}_t^4 - T\hat{\sigma}^4)/4 \end{bmatrix}$$

Inverting the Hessian, we get $\hat{\sigma}^2/T$ as its element *1,1*. Inverting the outer product, we get for its element *1,1*

$$(48) \qquad [(\sum_t \hat{u}_t^4 - T\hat{\sigma}^4)/4] \; / \; [T\hat{\sigma}^{-2}(\sum_t \hat{u}_t^4 - T\hat{\sigma}^4)/4 - (\sum_t \hat{u}_t^3/2)^2].$$

If the sum over time of $\hat{u}_t^3$ is zero, we get $\hat{\sigma}^2/T$ as from the Hessian; if this sum is not zero (as it can be for the small sample residuals), since we are subtracting its square at the denominator, we get a value which is smaller than $\hat{\sigma}^2/T$.

Summarizing, the element *1,1* of the inverse of the outer product is always greater or equal to the corresponding element of the inverted Hessian. In other words, for the variance of $\hat{\alpha}$, we always get from the outer product an estimate which is greater or equal to the estimate computed from the Hessian.

The experimental results displayed in section 5 suggest that something similar should hold also for systems of simultaneous equations. Although the sign of the inequality between variances does not occur in all the Monte Carlo replications, it occurs in a very high number of cases (from 90% to 100%, depending on the model and on the coefficient inside the model). Reminding that the random variable whose distribution is asymptotically approximated by χ^2 is obtained inverting the estimated covariance matrix, we get for such a variable smaller values when using the outer product, and the corresponding curve (3) always appears on the left of the Hessian curve (2).

Table 2

Multiplier-accelerator model

$$y_1 = a_1 y_3 + a_2 y_{1,t-1} + a_3 + u_1$$
$$y_2 = a_4 (y_3 - y_{3,t-1}) + a_5 y_{2,t-1} + a_6 + u_2$$
$$y_3 = y_1 + y_2 + x_1$$

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).

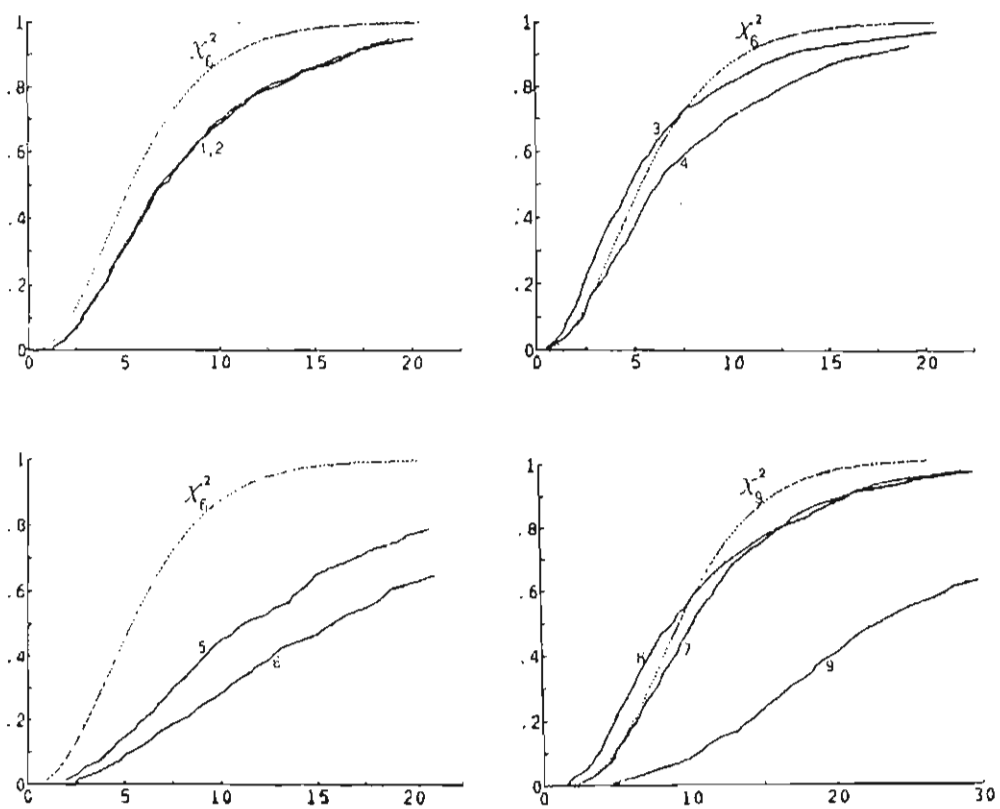

Fig.1. Multiplier-accelerator model. $T=20$; $n=6$; $n+m(m+1)/2=9$.

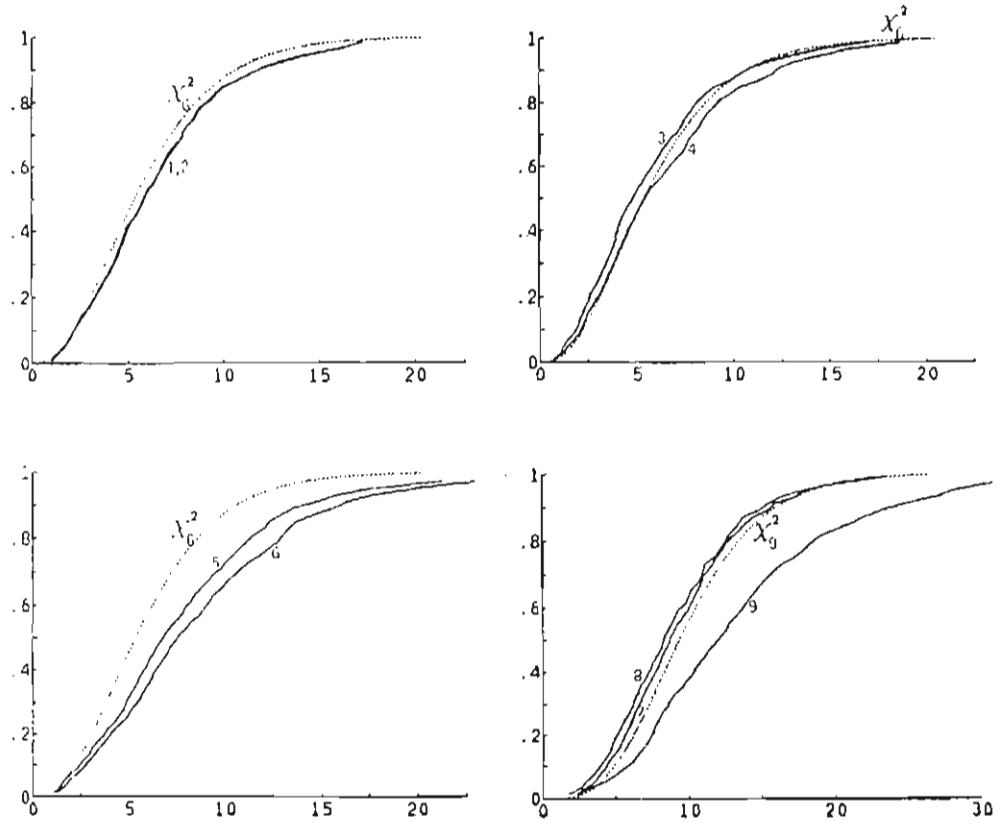

Fig.2. Multiplier-accelerator model. $T=50$; $n=6$; $n+m(m+1)/2=9$.

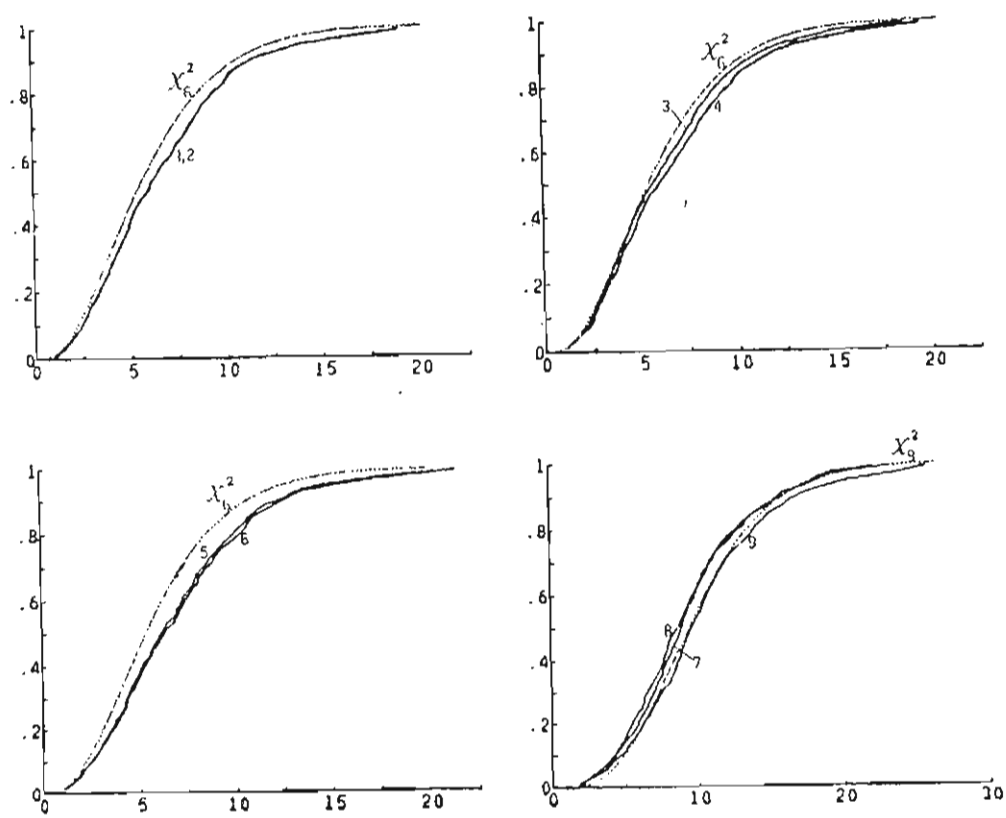

Fig.3. Multiplier-accelerator model. $T=200$; $n=6$; $n+m(m+1)/2=9$.

Table 3

Klein-I model

$$y_1 = a_1 + a_2 y_5 + a_3 y_{5,t-1} + a_4(y_3 + x_1) + u_1$$
$$y_2 = a_5 + a_6 y_5 + a_7 y_{5,t-1} + a_8 y_{6,t-1} + u_2$$
$$y_3 = a_9 + a_{10}(y_4 + x_2 - x_1) + a_{11}(y_{4,t-1} + x_{2,t-1} - x_{1,t-1}) + a_{12} x_3 + u_3$$
$$y_4 = y_1 + y_2 + x_4 - x_2$$
$$y_5 = y_4 - y_3 - x_1$$
$$y_6 = y_{6,t-1} + y_2$$

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).

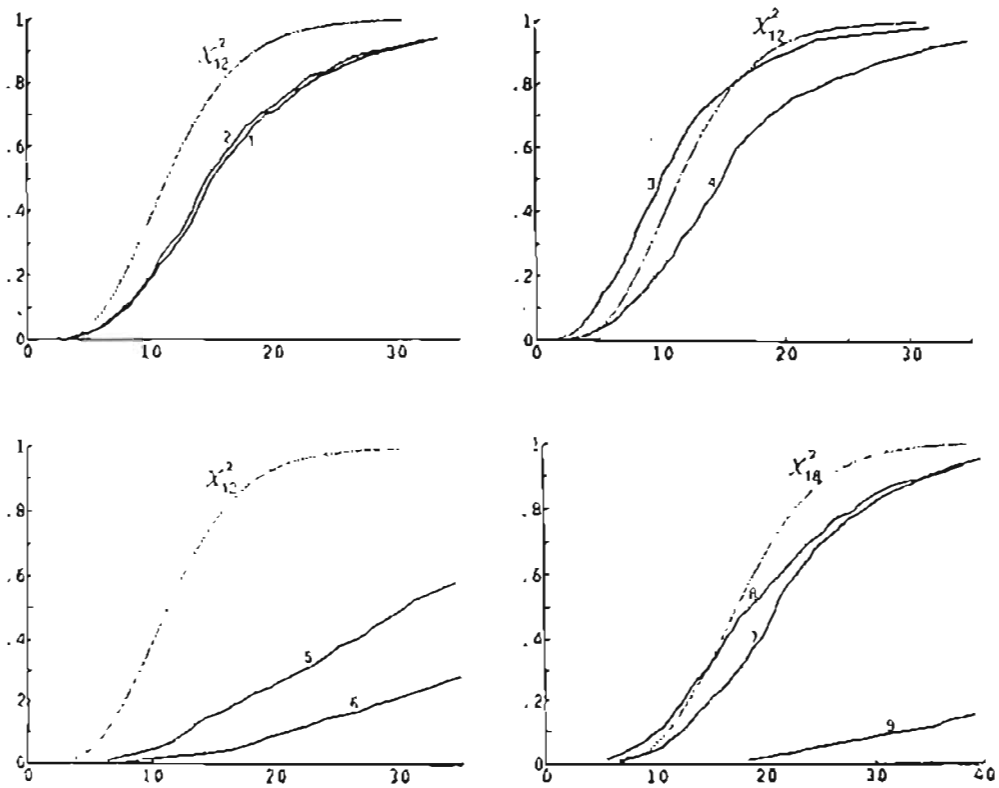

Fig.4. Klein-I model. $T=30$; $n=12$; $n+m(m+1)/2=18$.

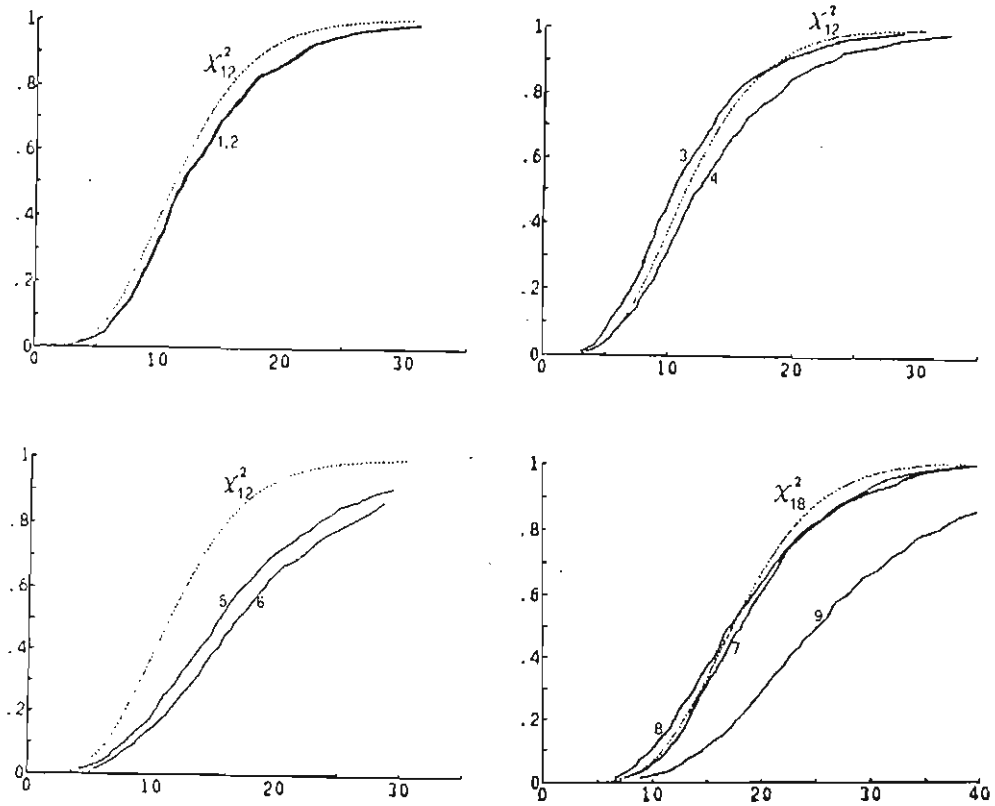

Fig.5. Klein-I model. $T=100$; $n=12$; $n+m(m+1)/2=18$.

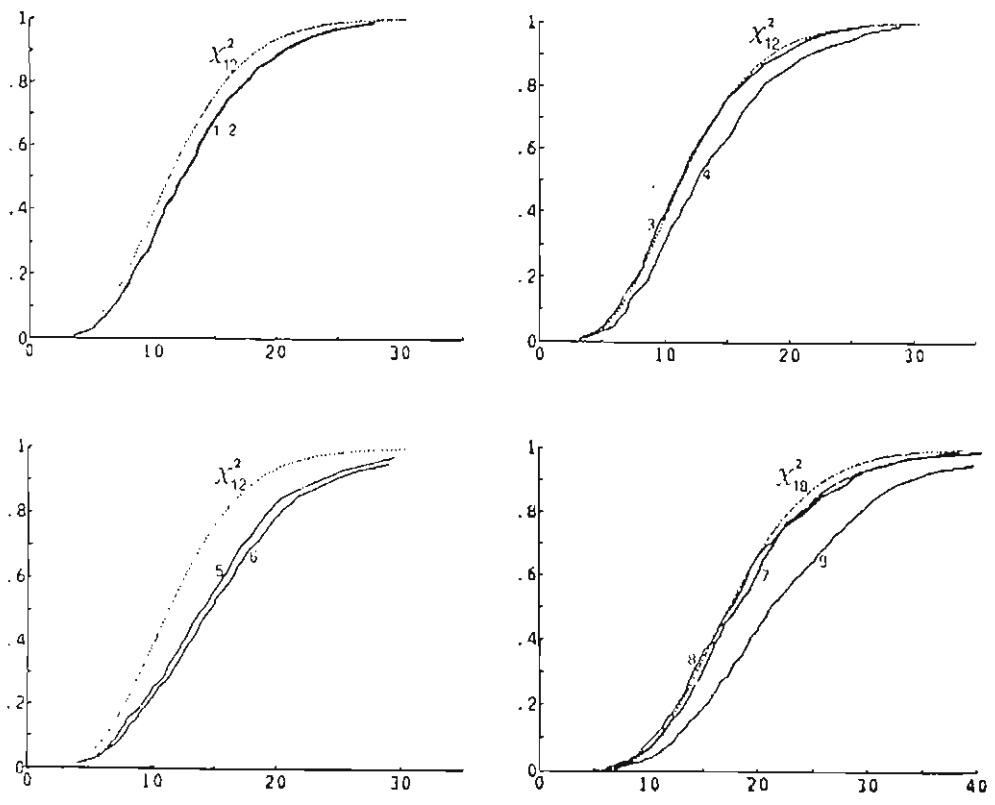

Fig.6. Klein-I model. $T=200$; $n=12$; $n+m(m+1)/2=18$.

Table 4

A model for consumption and price of food

$$y_1 = a_1 y_2 + a_2 + a_3 x_1 + u_1$$
$$y_1 = a_4 y_2 + a_5 + a_6 x_2 + a_7 x_3 + u_2$$

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).

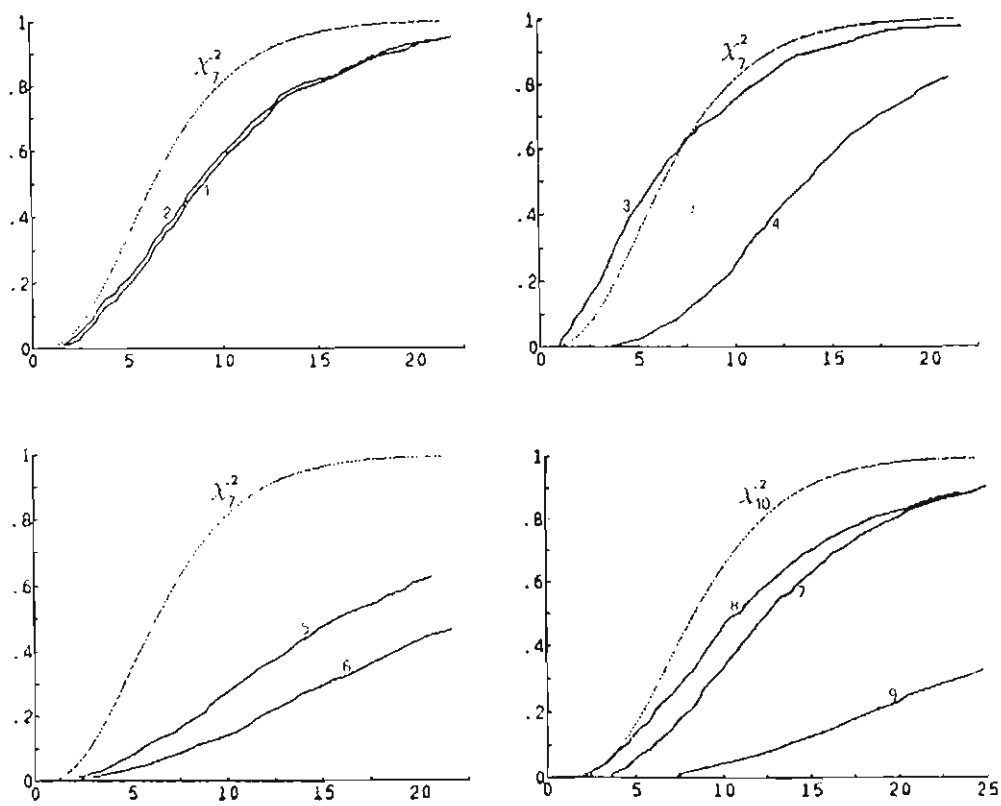

Fig.7. Food consumption and price model. $T=20$; $n=7$; $n+m(m+1)/2=10$.

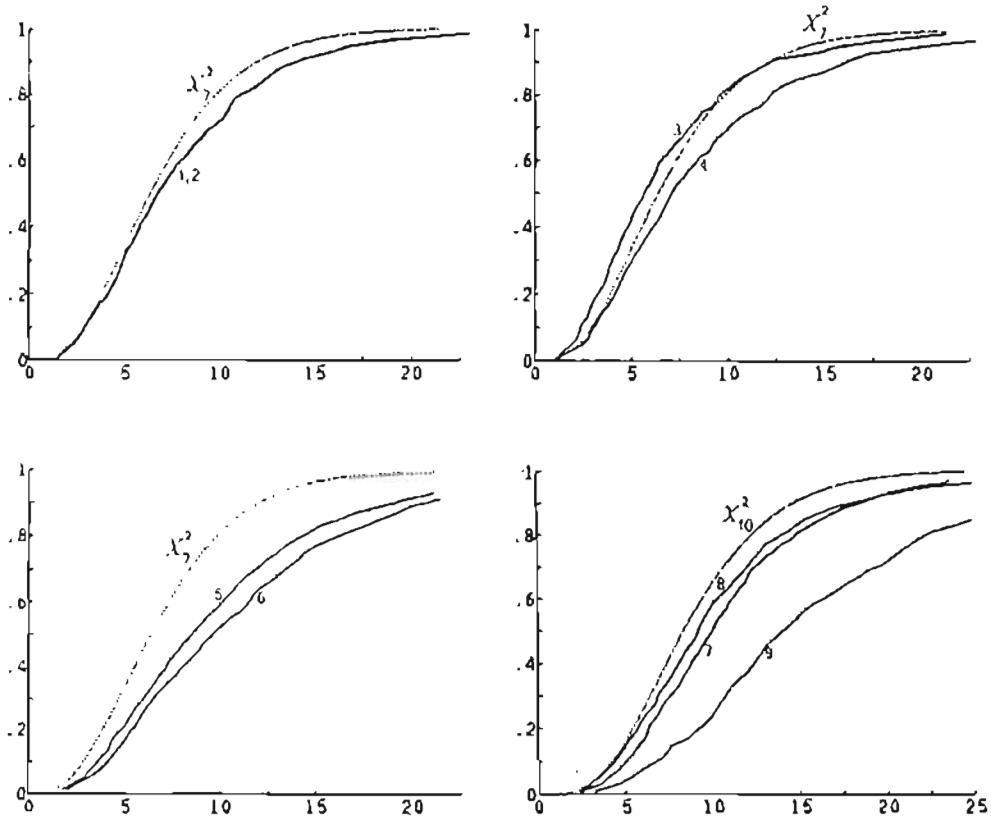

Fig.8. Food consumption and price model. $T=50$; $n=7$; $n+m(m+1)/2=10$.

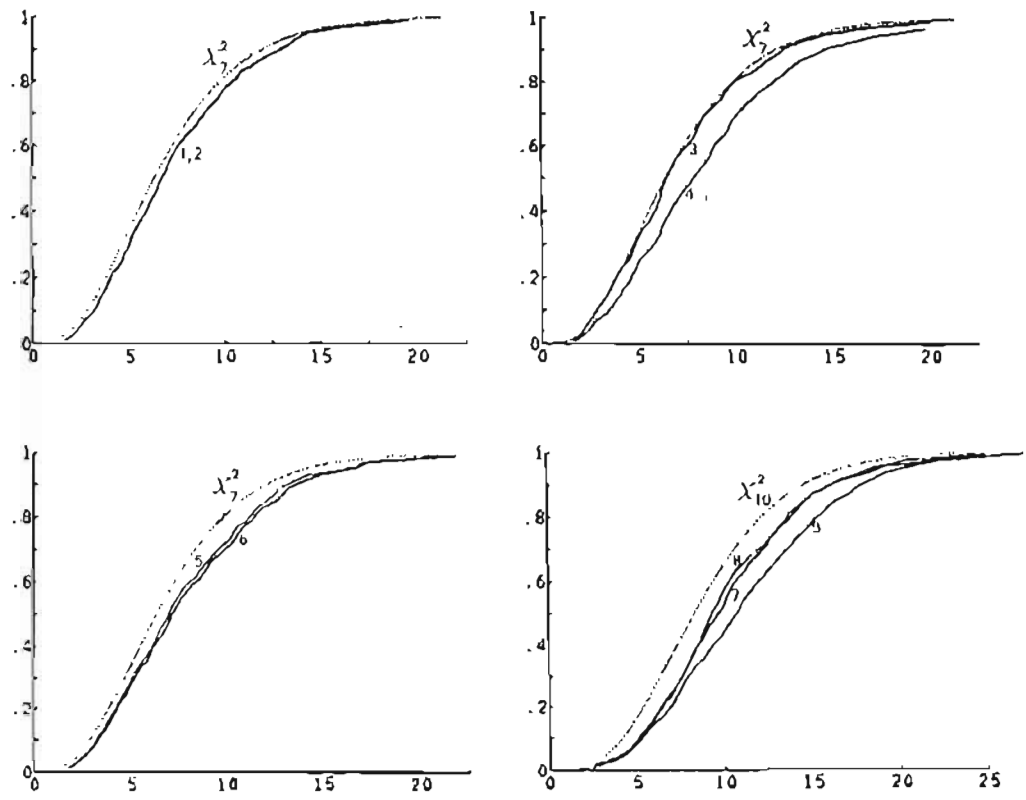

Fig.9. Food consumption and price model. $T=200$; $n=7$; $n+m(m+1)/2=10$.

Table 5

A model for the Italian economy

$$y_1 = a_1 + a_2(y_4+x_1+x_2) + a_3(y_6+x_3) + a_4(y_{6,t-1}+x_{3,t-1}) + u_1$$
$$y_2 = a_5 + a_6 y_{6,t-1} + a_7 y_5 + a_8 y_{2,t-1} + u_2$$
$$y_3 = a_9 + a_{10}(y_1+y_2) + a_{11} x_4 + u_3$$
$$y_4 = a_{12} + a_{13}(y_4+y_6) + a_{14} y_5 + a_{15} x_5 + a_{16} y_{4,t-1} + u_4$$
$$y_5 = a_{17} + a_{18}(y_2+y_{2,t-1}+y_{2,t-2}) + a_{19}(y_{2,t-1}+2y_{2,t-2}) + u_5$$
$$y_6 = y_7 - y_4 - x_1 - x_3 - x_2$$
$$y_7 = y_1 + y_2 - y_3 + x_1 - x_6 + x_7$$

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).

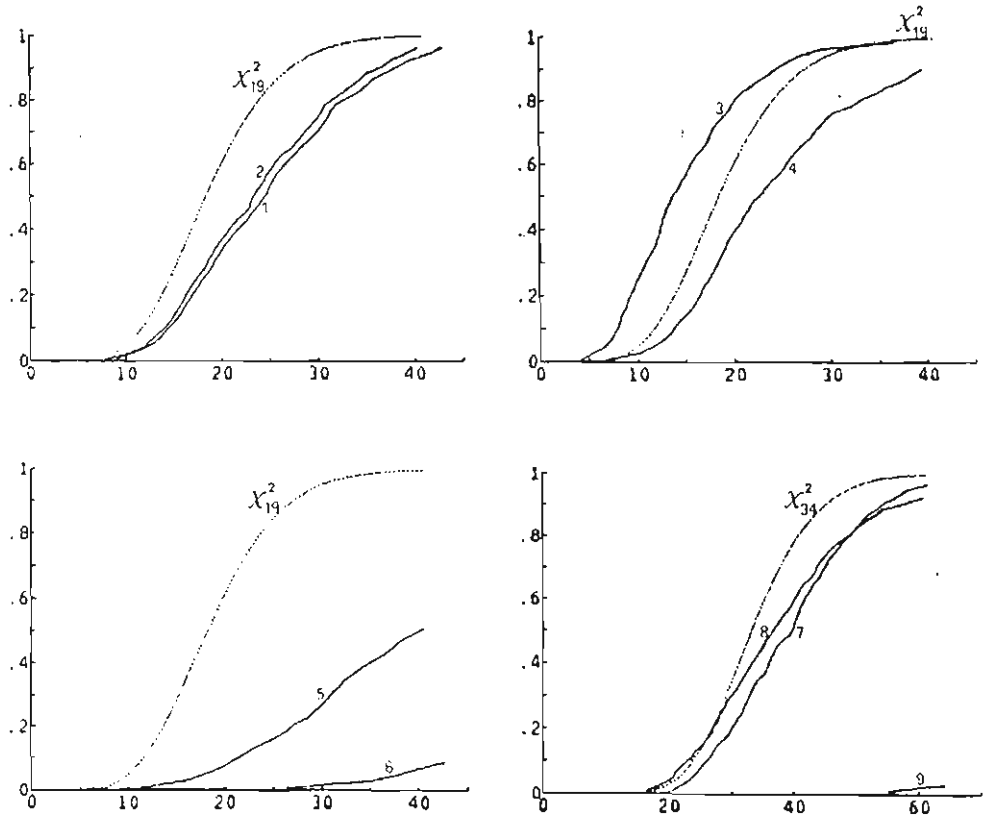

Fig. 10. Model for the Italian economy. $T=50$; $n=19$; $n+m(m+1)/2=34$.

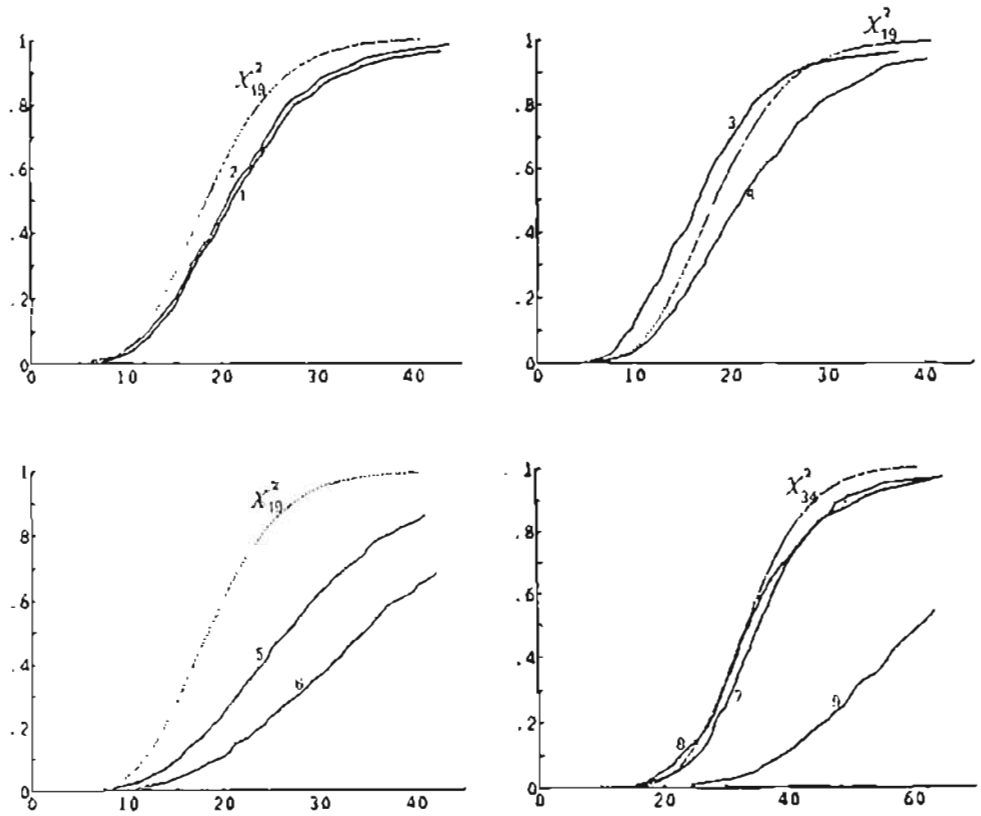

Fig.11. Model for the Italian economy. $T=100$; $n=19$; $n+m(m+1)/2=34$.

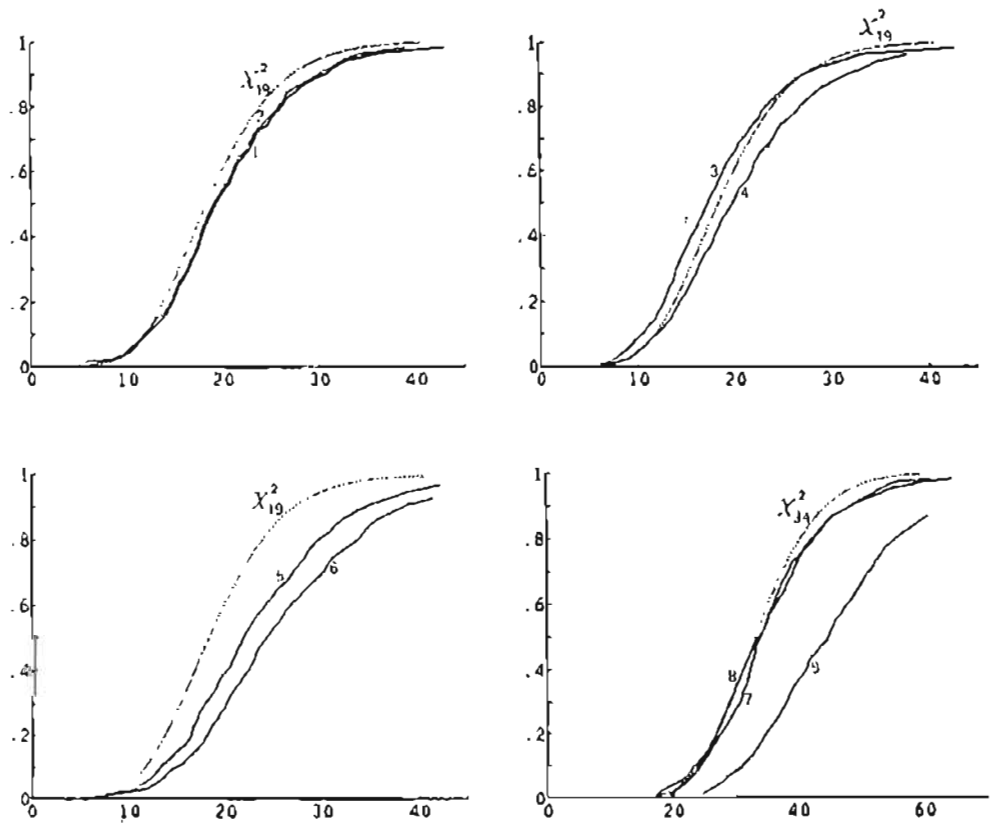

Fig.12. Model for the Italian economy. $T=200$; $n=19$; $n+m(m+1)/2=34$.

Table 6

Linearized 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).

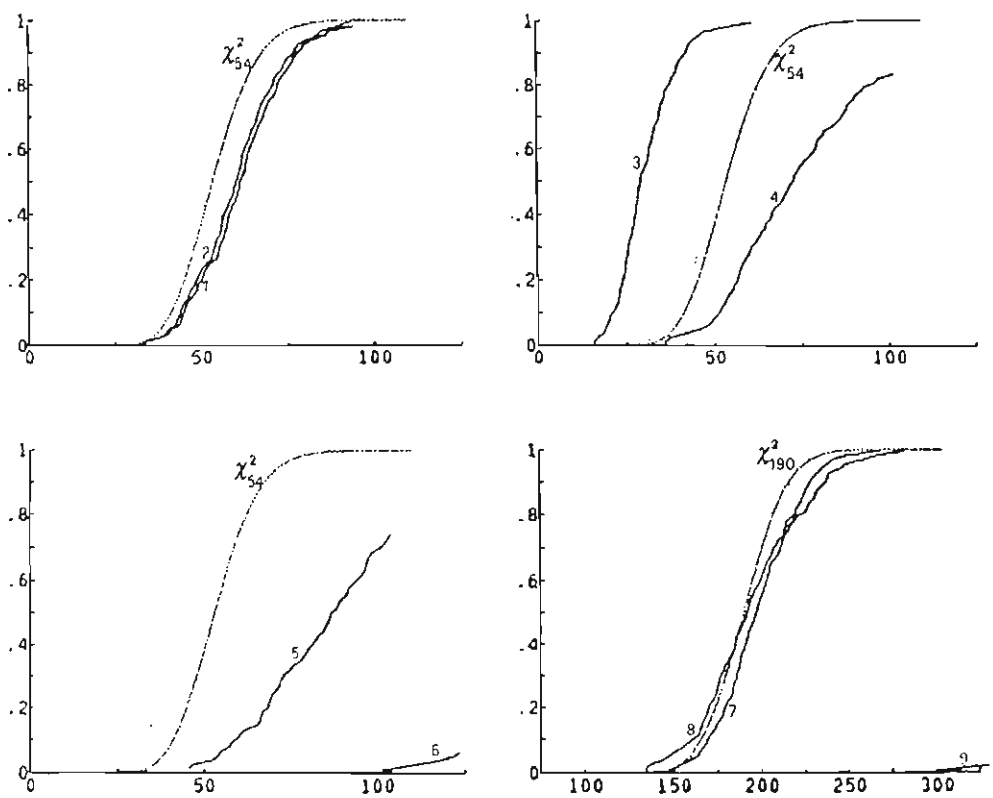

Fig. 13. Linearized Klein-Goldberger model. *T=300; n=54; n+m(m+1)/2=190.*

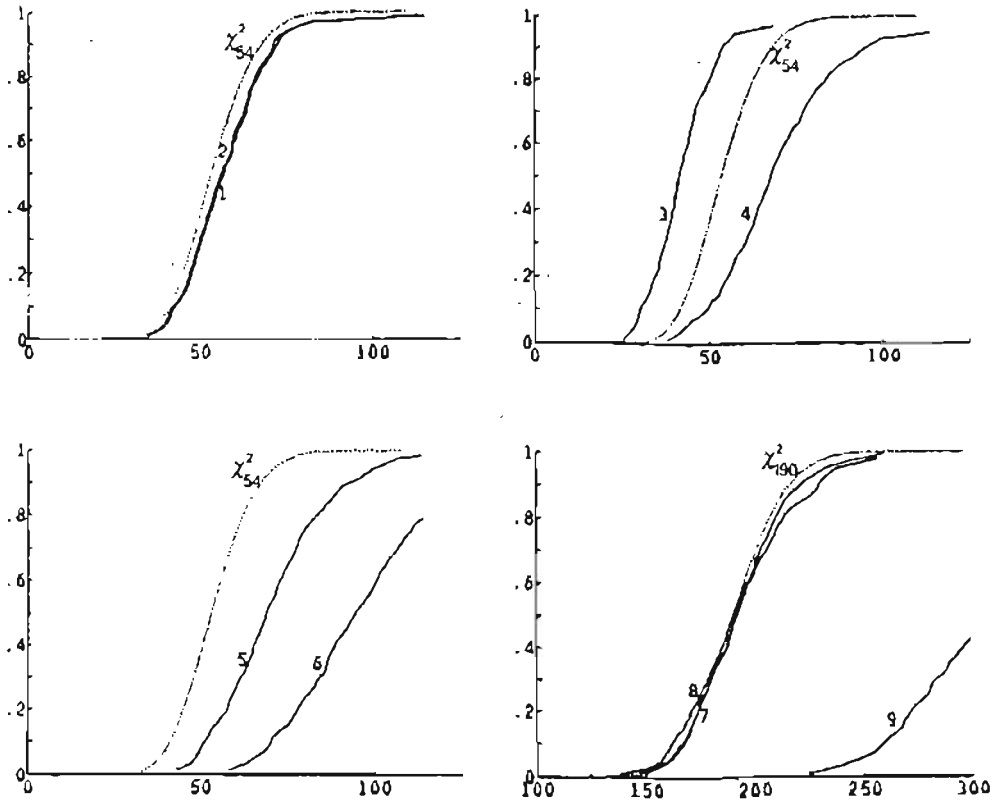

Fig.14. Linearized Klein-Goldberger model. $T=600$; $n=54$; $n+m(m+1)/2=190$.

Table 7

A simple Keynesian model

$$y_1 = a_1 y_2 + u_1$$
$$y_2 = y_1 + x_1$$

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.

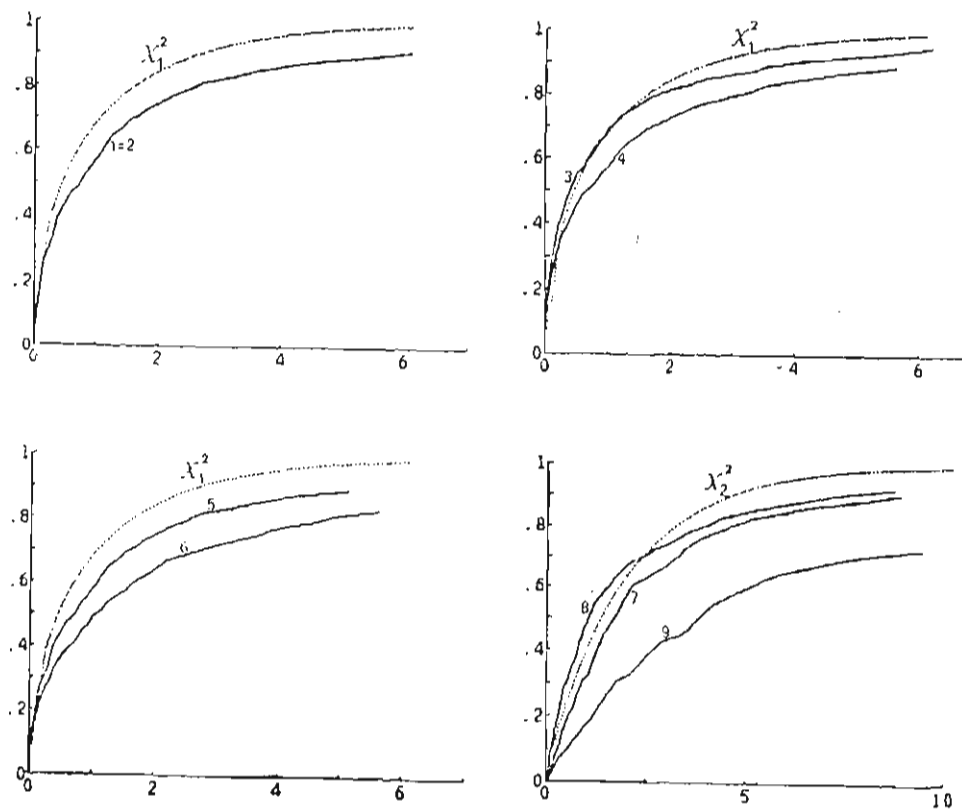

Fig. 15. Simple Keynesian model. $T=5$; $n=1$; $n+m(m+1)/2=2$.

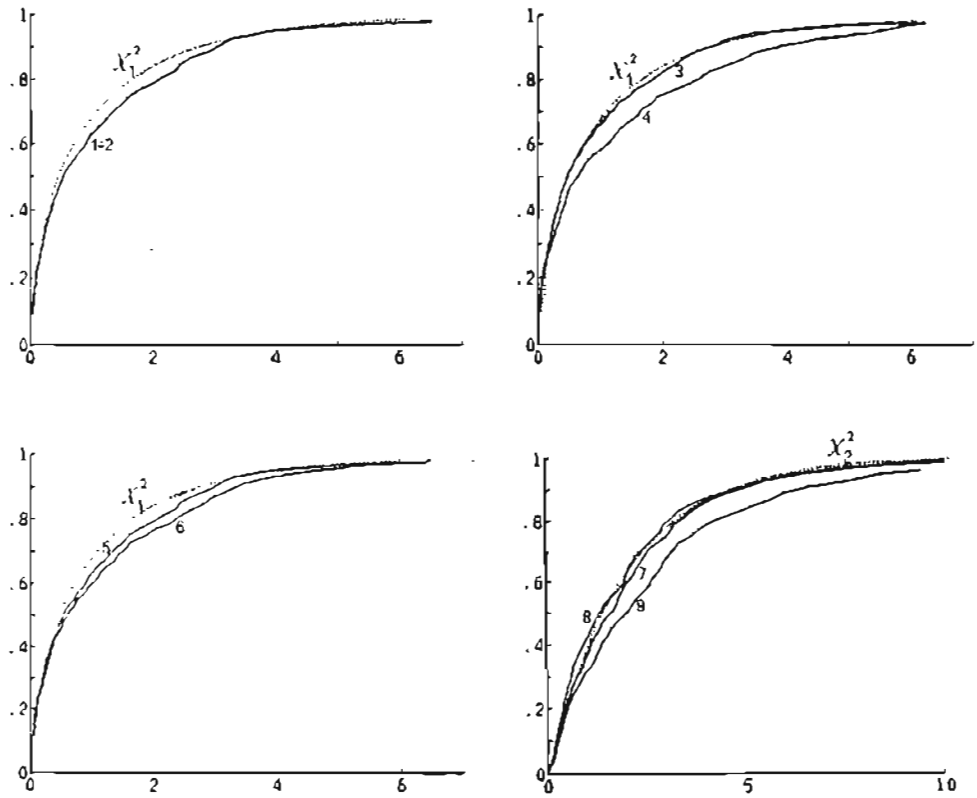

Fig.16. Simple Keynesian model. $T=20$; $n=1$; $n+m(m+1)/2=2$.

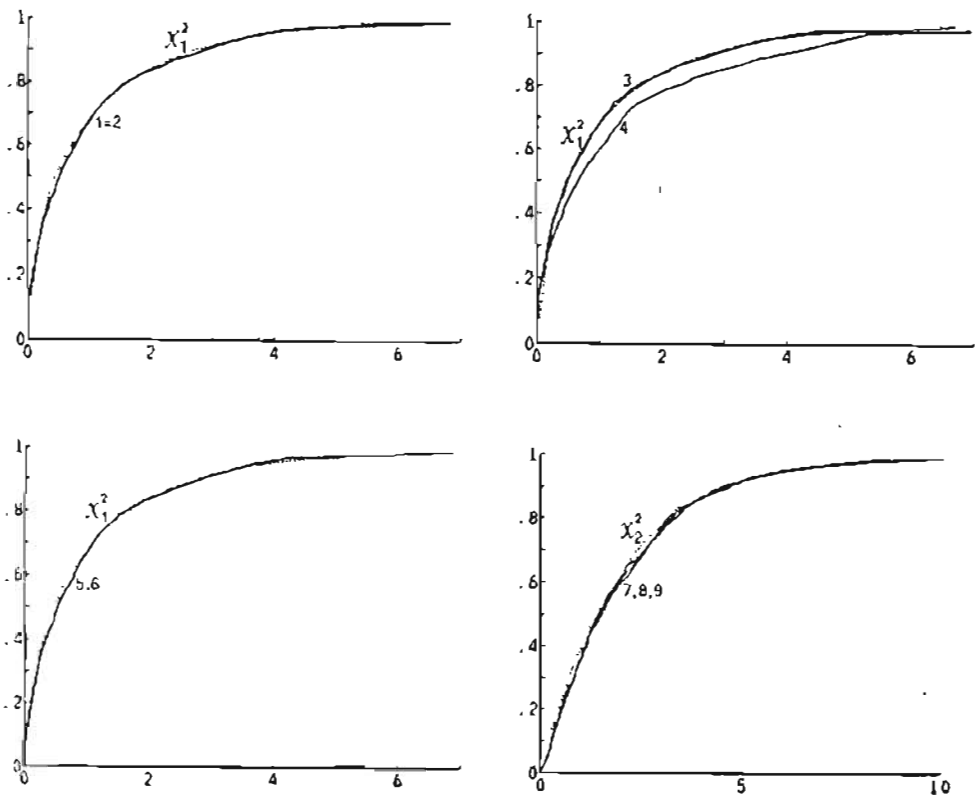

Fig.17. Simple Keynesian model. $T=300$; $n=1$; $n+m(m+1)/2=2$.

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.

Anderson,T.W., N.Kunitomo, and T.Sawa (1982), "Evaluation of the Distribution Function of the Limited Information Maximum Likelihood Estimator", *Econometrica* 50, 1009-1027.

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.

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. (1983), "Asymptotic Distribution of Power Spectra and Peak Frequencies in the Stochastic Response of Econometric Models", *Journal of Economic Dynamics and Control* 5, 235-247.

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.*

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.

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

Hall,A.R. (1983), "The Information Matrix Test for the Linear Model". University of Warwick: discussion paper presented at the *European Meeting of the Econometric Society, Pisa.*

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.

Morimune,K. (1983), "Approximate Distribution of k-Class Estimators when the Degree of Overidentifiability is Large Compared with the Sample Size", *Econometrica* 51, 821-841.

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.

Sitzia,B., and M.Tivegna (1975), "Un Modello Aggregato dell'Economia Italiana 1952-1971", in *Contributi alla Ricerca Economica* No.4. Roma: Banca d'Italia, 195-223.

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