AUGUST 1983

Revised JUNE 1984

N° 8424

A GENERAL APPROACH OF SERIAL CORRELATION

C. GOURIEROUX*, A. MONFORT**, A. TROGNON***

* Université Paris IX -Dauphine et CEPREMAP.

** INSEE

*** ENSAE

UNE APPROCHE GENERALE DE L'AUTOCORRELATION

C. GOURIEROUX, A. MONFORT, A. TROGNON

RESUME

Cet article étudie les problèmes de tests et d'estimation en présence d'autocorrélation. Le modèle général considéré recouvre des modèles très divers comme les modèles à équations simultanées non linéaires, les modèles probit, les modèles Tobit, les modèles de déséquilibre ; les modèles frontières... Dans ce contexte, il est montré que le test du score peut être complètement explicité et que le test obtenu généralise le test classique de Durbin et Watson ; par ailleurs il est établi que l'estimateur du maximum de vraisemblance est robuste vis à vis de l'autocorrélation.

A GENERAL APPROACH OF SERIAL CORRELATION

C. GOURIEROUX, A. MONFORT, A. TROGNON

ABSTRACT

In this article the testing and estimation problems are discussed in the case of serial correlation. Various models are particular cases of the general framework considered : the non-linear simultaneous equations models, the probit models, the tobit models, the desequilibrium models, the frontiers models... In this context, it is shown that the score test can be explicited and that the statistic obtained is a generalization of the Durbin-Watson'one ; on the other hand the maximum likelihood estimation procedure is shown to be robust with respect to serial correlation.

An important improvement of the econometric methods for time series data has been the introduction of the serial correlation. Since the well-known DURBIN and WATSON'work (1951) solving the problem of testing the serial correlation in the linear model, many studies have proposed testing and estimation procedures for various kind of models with serial correlation. However the solutions proposed in this literature are not yet unified and, moreover, they do not cover the important case of the so-called "limited dependent variables models", embedding qualitative response models, tobit models, disequilibrium models...

In this paper, we propose a general approach of the problem of testing serial correlation. This approach is a score type approach and it rests upon a very useful result, giving the form of the score vector in a transformed model, which is surprisingly ignored in the econometric literature. Our results are valid for a large class of models, including the limited-dependent variable models, and they naturally introduce the notion of generalized residuals which will be obviously useful for many other problems. Moreover, we show that the robustness property of the maximum likelihood method with respect to the serial correlation, proved by ROBINSON (1982) for the Tobit model, remains valid for a large class of models.

In section 2, we set up the class of models considered in the main part of the paper ; in section 3 we derive and discuss the likelihood function of this general model ; the problem of testing serial correlation is solved in section 4 and the estimation problem is considered in section 5 ; section 6 is devoted to various extensions.

2. THE MODEL.

2.a. Description of the model.

The observed endogenous variables are deduced from latent endogenous variables satisfying a nonlinear simultaneous equations model with autocorrelated errors.

Let y_t^* be the p-dimensional vector of the latent variables and let $\mathcal{Y}^* \subset \mathbb{R}^p$ be the range of y_t^* . These latent variables are such that :

$$(1) \qquad h(y_t^*, x_t; b) = u_t$$

where x_t is a q-dimensional vector of exogenous variables, b is a k-dimensional vector of parameters and u_t is the p-dimensional vector of disturbances. $h(., x_t; b)$ is assumed to be a one to one function from $\mathcal{Y}^*$ onto $\mathbb{R}^p$; it is thus possible to derive from equation (1) a well-defined reduced form.

In this section and in the two following ones, we consider the case of an autoregressive error of order one :

(2) $$u_t = R\, u_{t-1} + \varepsilon_t$$

where R is a $p \times p$ non-singular matrix, whose eigen-values are inside the unit circle and where $\varepsilon = (\varepsilon_t , t \in \mathbb{Z})$ is a white noise process with zero mean and with a regular variance-covariance matrix Ω :

(3) $$\varepsilon_t \quad \text{i.i.d} \quad , \quad \varepsilon_t \sim N(o,\Omega)$$

The m observed endogenous variables, denoted by y_t , are deduced from y_t^* by a known mapping g from $\mathcal{Y}^* \subset \mathbb{R}^p$ onto $\mathcal{Y} \subset \mathbb{R}^m$:

(4) $$y_t = g(y_t^*)$$

As it will be seen from the following examples, this formulation includes as special cases a great number of important econometric models .

2.b. Some examples.

Example 1 : When the observed variables coïncide with the latent ones, i.e when g is the identity function, the previous formulation contains the cases of the usual linear model, or of linear simultaneous equations models, but it also contains nonlinear unidimensional models such as

$$\frac{y_t^{\lambda} - 1}{\lambda} = x_t' c + u_t$$

or nonlinear simultaneous equations models such as :

$$\begin{cases} \text{Log } y_{1t} = a_1 y_{2t} + x_{1t}' b_1 + u_{1t} \\ y_{2t} = a_2 y_{1t} + x_{2t}' b_2 + u_{2t} \end{cases} \quad \text{with : } a_1 a_2 < o$$

Example 2 : Another kind of applications concerns limited dependent variable models, such as probit, tobit, frontier or disequilibrium models. The serial correlation in such models seems to be an important point and has been discussed by MAC RAE KENNAN (1982), GOURIEROUX-MONFORT-TROGNON (1984) for probit models, by DAGENAIS (1982), ROBINSON (1982), CHESHER-IRISH (1984) for tobit models, by ARTUS-LAROQUE-MICHEL (1984) for disequilibrium models.

Some of these models will be used as illustrations in further parts of the paper. In particular we shall consider the probit model which is defined by :

(5) $$y_t^* = x_t b + u_t \quad ; \quad u_t = \rho\, u_{t-1} + \varepsilon_t \quad , \quad \varepsilon_t \sim N(o,1)$$

$$y_t = 1\!\!1_{y_t^* > o} = \begin{cases} 1 & \text{if} \quad y_t^* > o \\ o & \text{otherwise} \end{cases}$$

(The variance of ε_t is taken equal to 1 in order to solve the usual identification problem) ; we shall also use a disequilibrium model defined by :

(6) $$\begin{cases} y_{1t}^* = x_{1t}' b_1 + u_{1t} \\ y_{2t}^* = x_{2t}' b_2 + u_{2t} \end{cases} \qquad u_t = R\, u_{t-1} + \varepsilon_t \ , \ \varepsilon_t \sim N(o,\Omega)$$

$$y_t = \text{Min}\,(y_{1t}^*, y_{2t}^*) \ .$$

3. DERIVATION OF THE LIKELIHOOD FUNCTION.

To build up the likelihood function, we successively consider the density function of $y^* = (y_1^* \ldots y_T^*)'$ and the density function of the observables $y = (y_1 \ldots y_T)'$.

The first one is easily obtained if it is possible to apply a Jacobian formula. We assume that the regularity conditions for this formula to be valid are satisfied. This is the case if function h is differentiable and has a differentiable inverse ; this is also the case for some piecewise differentiable functions such as the function which appears in multimarket disequilibrium models (see GOURIEROUX-LAFFONT-MONFORT (1980)).

3.a. Density function of the errors.

The p.d.f. of $(u_1,\dots,u_T)$ is :

$$\psi(u_1,\dots,u_T) = \frac{1}{(2\pi)^{p/2}\sqrt{\det Q}} \exp -\frac{1}{2}\, u_1' Q^{-1} u_1$$

$$\prod_{t=2}^{T} \frac{1}{(2\pi)^{p/2}\sqrt{\det \Omega}} \exp -\frac{1}{2}\,(u_t - Ru_{t-1})'\Omega^{-1}(u_t - Ru_{t-1})$$

where : $Q = \sum_{i=o}^{\infty} R^i \,\Omega\, R^{i'}$ is the marginal variance-covariance matrix of u_t .

The invertibility of Q is a direct consequence of the invertibility of Ω and R .

3.b. Density function of the latent variables.

The density function of $(y_1^*,\dots,y_T^*)$ with respect to the Lebesgue measure on $\mathcal{Y}^{*T}$, denoted by μ , is obtained by the Jacobian formula :

$\ell^*(y_1^*,\ldots,y_T^*;b,\Omega,R)$

$$= J_1(y_1^*;b)\,\frac{1}{(2\Pi)^{p/2}\sqrt{\det Q}}\,\exp -\frac{1}{2}\,h(y_1^*,x_1;b)'\,Q^{-1}\,h(y_1^*,x_1;b)$$

$$\prod_{t=2}^{T} J_t(y_t^*;b)\,\frac{1}{(2\Pi)^{p/2}\sqrt{\det\Omega}}\,\exp -\frac{1}{2}\,\{h(y_t^*,x_t;b) - R\,h(y_{t-1}^*,x_{t-1};b)\}'$$

$$\Omega^{-1}\,\{h(y_t^*,x_t;b) - R\,h(y_{t-1}^*,x_{t-1};b)\} \quad .$$

where $J_t(.;b)$ is the jacobian determinant of the mapping $h(.,x_t;b)$.

When the hypothesis of noncorrelation $Ho : (R=o)$ is satisfied this expression becomes :

$$\ell^*(y_1^*,\ldots,y_T^*;b,\Omega,o) = \prod_{t=1}^{T} \ell_o^*(y_t^*;b,\Omega)$$

where : $$\ell_o^*(y_t^*;b,\Omega) = J_t(y_t^*;b)\,\frac{1}{(2\Pi)^{p/2}\sqrt{\det\Omega}}$$

$$\exp -\frac{1}{2}\,h(y_t^*,x_t;b)'\,\Omega^{-1}\,h(y_t^*,x_t;b)$$

is the marginal density function of y_t^* under Ho .

3.c. <u>Density function of the observable endogenous variables.</u>

In order to easily derive this function, we first establish a lemma, which shows that the density function ℓ of the observable variables can be deduced from the density function ℓ^* of the latent variables by integration with respect to a measure which does not depend on the parameters.

Let us consider a particular value $\tilde{\theta}$ of the parameter $\theta = (b,\Omega,R)$; the measure μ can be replaced by the probability distribution :

$$P^*_{\tilde{\theta}} = \ell^*(y^*;\tilde{\theta}).\mu = \ell^*(y^*_1 \ldots y^*_T;\tilde{b},\tilde{\Omega},\tilde{R}).\mu$$

The density function of $P^*_\theta = \ell^*(y^*;\theta).\mu$ with respect to $P^*_{\tilde{\theta}}$ is : $f^*_{\tilde{\theta}}(y^*;\theta) = \dfrac{\ell^*(y^*;\theta)}{\ell^*(y^*;\tilde{\theta})}$.

For each value θ of the parameter, the associated probability distribution of the observables y is denoted by P_θ .

<u>Lemma</u> : The probability distribution P_θ has a density function with respect to $P_{\tilde{\theta}}$. This density function, denoted by $f_{\tilde{\theta}}(y;\theta)$, is given by :

$$f_{\tilde{\theta}}(Y;\theta) = E_{\tilde{\theta}}[f^*_{\tilde{\theta}}(Y^*;\theta)/Y]$$

where $E_{\tilde{\theta}}(\ /Y)$ is the conditional expectation associated with the value $\tilde{\theta}$ of the parameter.

Proof : For any positive measurable function φ , we have :

$$\int_{y^T} \varphi(y)\, dP_\Theta(y)$$

$$= \int_{(y^*)^T} \varphi[g(y^*)]\, dP^*_\Theta\,(y^*) \qquad \text{with} \quad g(y^*) = [g(y^*_1),\dots,g(y^*_T)]$$

$$= \int_{(y^*)^T} \varphi[g(y^*)]\, f^*_{\underset{\Theta}{\sim}}(y^*;\theta)\, dP^*_{\underset{\Theta}{\sim}}(y^*)$$

$$= \int_{y^T} \int_{(y^*)^T} \varphi[g(y^*)]\, f^*_{\underset{\Theta}{\sim}}(y^*;\theta)\, dP^*_{\underset{\Theta}{\sim}}(y^*/y)\, dP_{\underset{\Theta}{\sim}}(y)$$

(where $dP^*_{\underset{\Theta}{\sim}}(y^*/y)$ is the conditional distribution of Y^* given Y associated with $\tilde{\Theta}$)

$$= \int_{y^T} \varphi(y) \left[\int_{(y^*)^T} f^*_{\underset{\Theta}{\sim}}(y^*;\theta)\, dP^*_{\underset{\Theta}{\sim}}(y^*/y)\right] dP_{\underset{\Theta}{\sim}}(y)$$

$$= \int_{y^T} \varphi(y)\, E_{\underset{\Theta}{\sim}}[f^*_{\underset{\Theta}{\sim}}(Y^*;\theta)/Y=y]\;\; dP_{\underset{\Theta}{\sim}}(y)$$

Therefore : $P_\Theta = f_{\underset{\Theta}{\sim}}(y;\theta)\, .\, P_{\underset{\Theta}{\sim}}$

with : $f_{\underset{\Theta}{\sim}}(Y;\theta) = E_{\underset{\Theta}{\sim}}[f^*_{\underset{\Theta}{\sim}}(Y^*;\theta)/Y]$

Q.E.D.

The density function $f_{\underset{\Theta}{\sim}}(y;\theta)$ is obtained by integrating $f^*_{\underset{\Theta}{\sim}}(y^*;\theta)$ with respect to the conditional distribution $P^*_{\underset{\Theta}{\sim}}(dy^*/y)$ or, equivalently by integrating the initial density function $\ell^*(y^*;\theta)$ with respect to the measure :

$\nu_{\tilde{\theta}}(dy^*/y) = \frac{1}{\ell^*(y^*;\tilde{\theta})} \cdot P^*_{\tilde{\theta}}(dy^*/y)$ which is independent from the parameter θ.

Therefore the probability distribution of the observable endogenous variables has a density function ℓ which can be written as :

$$(7) \qquad \ell(y;\theta) = \int_{(y^*)^T} \ell^*(y^*;\theta)\ \nu(dy^*/y)$$

where ν is a fixed measure.

Remark : It may be interesting to choose a value $\tilde{\theta}$ for which the latent variables are uncorrelated. In this case the measure $\nu_{\tilde{\theta}}$ is a product of measures :

$$\nu_{\tilde{\theta}}(dy^*/y) = \prod_{t=1}^{T} \frac{1}{\ell^*_o(y^*_t,\tilde{\theta})} P^*_{o,\tilde{\theta}}(dy^*_t/y_t) \quad .$$

3.d. Some difficulties.

The previous result can be illustrated by considering a probit model. In this case the density function of the observables may be chosen as :

$$\ell(y,\theta) = \ell(y_1,\dots,y_T;b,\rho)$$

$$= \int_{y^*_1 \gtrless o} \dots \int_{y^*_T \gtrless o} \ell^*(y^*_1,\dots,y^*_T;b,\rho)\ dy^*_1 \dots dy^*_T$$

where the symbol $y_t^* \gtrless o$ means that the set of integration is $y_t^* > o$ if $y_t = 1$, is $y_t^* < o$ if $y_t = o$.

In general, the density function of the observables takes the form of a multi-dimensional integral whose dimension is the number of observations T. This creates serious problems for the maximum likelihood estimators of the parameters and for the test procedures of hypotheses such as $Ho : \rho = o$ (non correlation) :

i) The numerical maximization of the likelihood function with respect to b, Ω, R is likely to be very difficult (in the general case).

ii) The asymptotic tests of Ho, based on the unconstrained M L estimators, are also untractable. This remark is valid for the Wald test or for the likelihood ratio test.

iii) This difficulty does not arise with the score test, or the Lagrange multiplier test, which is based on the constrained M L estimator. In effect under the null, $\ell^*(y_1^*, \ldots, y_T^*; b, \Omega, R)$ is a product $\prod_{t=1}^{T} \ell_o^*(y_t^*; b, \Omega)$ and, from the remark, the multidimensional integral reduces to a product of integrals whose dimension is equal to p and, therefore, is independent from the number of observations.

iv) However, even if the score statistic can be numerically evaluated, its asymptotic properties are not known. In effect they are usually derived by using the asymptotic properties of the constrained and of the unconstrained M L estimators, but the form of the likelihood function does not satisfy the usual regularity conditions ensuring the consistency and the asymptotic normality of the unconstrained M L estimators.

In the following sections we develop several testing and estimation procedures applicable in nonlinear models with serial correlation. These procedures are based on asymptotic results and we assume that the classical regularity conditions, ensuring consistency and asymptotic normality of M-estimators are satisfied. (See BURGUETE -GALLANT-SOUZA (1982)).

4. SCORE TEST OF THE NON CORRELATION HYPOTHESIS.

4.a. Links between observable and latent scores.

From the previous lemma, we know that the density function of y may be chosen as :

$$\ell(y;\theta) = \int_{(y^*)^T} \frac{\ell^*(y^*;\theta)}{\ell^*(y^*;\tilde{\theta})} P^*_{\tilde{\theta}}(dy^*/y)$$

Assuming that regularity conditions for commuting derivation and integration are satisfied, we deduce that :

$$\frac{\partial \log \ell(y;\theta)}{\partial \theta} = \frac{\int_{(y^*)^T} \frac{\partial}{\partial \theta} \ell^*(y^*;\theta) \frac{1}{\ell^*(y^*;\tilde{\theta})} P^*_{\tilde{\theta}}(dy^*/y)}{\int_{(y^*)^T} \frac{\ell^*(y^*;\theta)}{\ell^*(y^*;\tilde{\theta})} P^*_{\tilde{\theta}}(dy^*/y)}$$

Noting that the score is independent from the dominating measure, in particular independent from $\tilde{\theta}$, we obtain by setting $\tilde{\theta} = \theta$:

$$\frac{\partial \log \ell(y;\Theta)}{\partial \Theta} = \int_{(y^*)^T} \frac{\partial}{\partial \Theta} \ell^*(y^*;\Theta) \frac{1}{\ell^*(y^*;\Theta)} P^*_\Theta(dy^*/y)$$

$$(8) \quad \boxed{\frac{\partial \log \ell(y;\Theta)}{\partial \Theta} = E_\Theta\left[\frac{\partial \log \ell^*(Y^*;\Theta)}{\partial \Theta} / Y = y\right]}$$

The observable score is equal to the conditional expectation of the latent score given the observable endogenous variables.

4.b. Expression of the observable score.

The score test of the hypothesis $Ho : (R = o)$ is based on the score estimated under the null, i.e on :

$$\hat{\xi} = \left[\frac{\partial \log \ell(y;b,\Omega,R)}{\partial \text{ vec } R}\right]_{b=\hat{b}_{oT},\Omega=\hat{\Omega}_{oT},R=o}$$

where $\hat{b}_{oT}$ and $\hat{\Omega}_{oT}$ are the M L estimator of b and Ω constrained by Ho.

From (8), we deduce that :

$$\xi = \frac{\partial \log \ell(y;b,\Omega,o)}{\partial \text{ vec } R} = E_o \left[\frac{\partial \log \ell^*(Y^*;b,\Omega,o)}{\partial \text{ vec } R} / Y = y\right]$$

where $E_o(/Y=y)$ denotes the conditional expectation under the null.

We have :

$$\frac{\partial \log \ell^*(y^*;b,\Omega,o)}{\partial \text{ vec } R} = \left\{ \frac{\partial}{\partial \text{ vec}R} \quad [-\frac{1}{2} \log \det Q] \right.$$

$$+ \frac{\partial}{\partial \text{ vec}R} \quad [-\frac{1}{2} u_1' \; Q^{-1} \; u_1] \; + \; \sum_{t=2}^{T} \frac{\partial}{\partial \text{ vec}R} \; (- \frac{1}{2} \log \det \Omega \;\;)$$

$$\left. + \sum_{t=2}^{T} \frac{\partial}{\partial \text{vec}R} \; [- \frac{1}{2} (u_t - R \; u_{t-1})' \; \Omega^{-1} \; (u_t - R \; u_{t-1})] \right\}_{R=o}$$

$$= (\Omega^{-1} \otimes I) \sum_{t=2}^{T} u_t \otimes u_{t-1}$$

with $u_t = h(y_t^*, x_t, b)$.

Therefore :

$$\xi = \frac{\partial \log\ell(y;b,\Omega,o)}{\partial \text{ vec } R} = \underset{o}{E} \; [(\Omega^{-1} \otimes I) \sum_{t=2}^{T} (u_t \otimes u_{t-1}) \; / \; Y = y]$$

$$= \Omega^{-1} \otimes I \sum_{t=2}^{T} \underset{o}{E} \; (u_t \otimes u_{t-1} \; / \; Y = y)$$

Since, under the null, the vectors (u_t, y_t) are serially independent, we obtain :

$$\xi = (\Omega^{-1} \otimes I) \sum_{t=2}^{T} \underset{o}{E} \; (u_t / Y_t = y_t) \otimes \underset{o}{E} \; (u_{t-1} / Y_{t-1} = y_{t-1})$$

We denote by :

(9) $$\tilde{u}_t(y_t, x_t, b, \Omega) = \underset{o}{E} \; (h(Y_t^*, x_t, b) \; / \; Y_t = y_t) = \underset{o}{E} \; (u_t / Y_t = y_t)$$

the prediction of the disturbance u_t , evaluated under the null $\Theta = (b,\Omega,o)$, and by :

$$\text{(10)} \quad \boxed{\hat{\tilde{u}}_t = \tilde{u}_t(y_t, x_t, \hat{b}_{oT}, \hat{\Omega}_{oT})}$$

its estimation under the null.

The score statistic is thus given by :

$$\text{(11)} \quad \boxed{\hat{\xi} = (\hat{\Omega}_o^{-1} \otimes I) \sum_{t=2}^{T} \hat{\tilde{u}}_t \otimes \hat{\tilde{u}}_{t-1}}$$

This statistic is in one to one relationship with the empirical autocovariance of the "predicted" residuals, $\hat{\tilde{u}}_t$; these residuals will be called <u>generalized residuals</u>.

4.c. <u>Asymptotic distribution of $\hat{\xi}$ under the null</u>.

As noted in section 3.d , it is necessary to examine directly the asymptotic properties of $\hat{\xi}$ under the null.

Let us assume that the true value of the parameter is $\Theta_o = (b_o,\Omega_o,o)$ and let us denote by : $\hat{\Theta}_{oT} = (\hat{b}_{oT},\hat{\Omega}_{oT},o)$ the constrained M L estimator of Θ . Under a set of classical regularity conditions, including the asymptotic identifiability of θ under the null (see BURGUETE-GALLANT-SOUZA (1982)), it is possible to prove the strong consistency of $\hat{\theta}_{oT}$ to θ_o and to replace the statistic $\hat{\xi}$ by its expansion in a neighbourhood of θ_o . More precisely, we have :

$$\frac{1}{\sqrt{T}} \sum_{t=2}^{T} \hat{\tilde{u}}_t \otimes \hat{\tilde{u}}_{t-1} = \frac{1}{\sqrt{T}} \sum_{t=2}^{T} E_{\hat{\theta}_{oT}} (u_t/y_t) \otimes E_{\hat{\theta}_{oT}} (u_{t-1}/y_{t-1})$$

$$= \frac{1}{\sqrt{T}} \sum_{t=2}^{T} E_{\theta_o}(u_t/y_t) \otimes E_{\theta_o}(u_{t-1}/y_{t-1})$$

$$+ \left[\frac{1}{T} \sum_{t=2}^{T} E_{\theta_o}(u_t/y_t) \otimes \frac{\partial}{\partial \theta'} E_{\theta_o}(u_{t-1}/y_{t-1})\right] \sqrt{T} (\hat{\theta}_{oT} - \theta_o)$$

$$+ \left[\frac{1}{T} \sum_{t=2}^{T} \frac{\partial}{\partial \theta'} E_{\theta_o}(u_t/y_t) \otimes E_{\theta_o}(u_{t-1}/y_{t-1})\right] \sqrt{T} (\hat{\theta}_{oT} - \theta_o)$$

$$+ o_p(1)$$

$$= \frac{1}{\sqrt{T}} \sum_{t=2}^{T} E_{\theta_o}(u_t/y_t) \otimes E_{\theta_o}(u_{t-1}/y_{t-1})$$

$$+ \frac{1}{T} \sum_{t=2}^{T} E_{\theta_o} \left[E_{\theta_o}(u_t/y_t) \otimes \frac{\partial}{\partial \theta'} E_{\theta_o}(u_{t-1}/y_{t-1})\right] \sqrt{T} (\hat{\theta}_{oT} - \theta_o)$$

$$+ \frac{1}{T} \sum_{t=2}^{T} E_{\theta_o} \left[\frac{\partial}{\partial \theta'} E_{\theta_o}(u_t/y_t) \otimes E_{\theta_o}(u_{t-1}/y_{t-1})\right] \sqrt{T} (\hat{\theta}_{oT} - \theta_o) + o_p(1)$$

Under the null $E_{\Theta_0}(u_t/y_t)$ and $\frac{\partial}{\partial\Theta} E_{\Theta_0}(u_{t-1}/y_{t-1})$ are independent variables ; this implies that the second and the third terms of the previous equations are equal to zero since, we have, for instance :

$$E_{\Theta_0}\left[E_{\Theta_0}(u_t/y_t) \otimes \frac{\partial}{\partial\Theta'} E_{\Theta_0}(u_{t-1}/y_{t-1})\right]$$

$$= E_{\Theta_0} u_t \otimes E_{\Theta_0}\left[\frac{\partial}{\partial\Theta'} E_{\Theta_0}(u_{t-1}/y_{t-1})\right] = 0$$

Therefore : $\frac{1}{\sqrt{T}} \sum_{t=2}^{T} \hat{\tilde{u}}_t \otimes \hat{\tilde{u}}_{t-1}$

$$= \frac{1}{\sqrt{T}} \sum_{t=2}^{T} E_{\Theta_0}(u_t/y_t) \otimes E_{\Theta_0}(u_{t-1}/y_{t-1}) + o_P(1)$$

Let us now consider the asymptotic behavior of the empirical mean $\frac{1}{T} \sum_{t=2}^{T} E_{\Theta_0}(u_t/y_t) \otimes E_{\Theta_0}(u_{t-1}/y_{t-1})$.

Under regularity assumptions required for the strong law of large numbers and central limit theorem we have :

i) $\frac{1}{T} \sum_{t=2}^{T} E_{\Theta_0}(u_t/y_t) \otimes E_{\Theta_0}(u_{t-1}/y_{t-1})$

converges almost surely to

$$\lim \frac{1}{T}\sum_{t=1}^{T} E_{\Theta_0}\left[E_{\Theta_0}(u_t/y_t) \otimes E_{\Theta_0}(u_{t-1}/y_{t-1})\right] = \text{plim} \frac{1}{T} \sum_{t=1}^{T} E_{\Theta_0} u_t \otimes E_{\Theta_0} u_{t-1} = 0$$

ii) $\frac{1}{\sqrt{T}} \sum_{t=2}^{T} E_{\Theta_o}(u_t/y_t) \otimes E_{\Theta_o}(u_{t-1}/y_{t-1})$ is asymptotically normal with zero mean and with a covariance matrix equal to :

$$\sum_{j=-\infty}^{\infty} \Gamma_j$$

with $\begin{cases} \Gamma_j = \text{plim} \frac{1}{T} \sum_{t=2}^{T-j} E_{\Theta_o}\left[(\tilde{u}_t \otimes \tilde{u}_{t-1})(\tilde{u}_{t+j} \otimes \tilde{u}_{t+j-1})'\right] , \quad j > o \\ \Gamma_{-j} = \Gamma_j' . \end{cases}$

where plim denotes the probability limit with respect to the exogenous variables.

It is easily seen that $\Gamma_j = o \ \forall \ j \neq o$ and therefore :

$$\sum_{j=-\infty}^{\infty} \Gamma_j = \Gamma_o = \text{plim} \frac{1}{T} \sum_{t=2}^{T} E_{\Theta_o}\left[(\tilde{u}_t \otimes \tilde{u}_{t-1})(\tilde{u}_t \otimes \tilde{u}_{t-1})'\right]$$

$$= \text{plim} \frac{1}{T} \sum_{t=2}^{T} E_{\Theta_o}(\tilde{u}_t \ \tilde{u}_t') \otimes E_{\Theta_o}(\tilde{u}_{t-1}\tilde{u}_{t-1}')$$

$$= \text{plim} \frac{1}{T} \sum_{t=2}^{T} (\tilde{u}_t\tilde{u}_t') \otimes (\tilde{u}_{t-1}\tilde{u}_{t-1}')$$

Finally, the score statistic $\hat{\xi}$ is such that, under the null :

$$\frac{1}{T} \hat{\xi} \xrightarrow[T \to \infty]{a.s} o$$

$$\frac{1}{\sqrt{T}} \hat{\xi} \xrightarrow[T\to\infty]{d} N[o,(\Omega_o^{-1} \otimes I) \ \text{plim} \frac{1}{T} \sum_{t=2}^{T} (\tilde{u}_t\tilde{u}_t') \otimes (\tilde{u}_{t-1}\tilde{u}_{t-1}') \ (\Omega_o^{-1} \otimes I)]$$

4.d. The score test.

A consistent estimator of the asymptotic variance of $\frac{1}{\sqrt{T}} \hat{\xi}$ under the null is such that :

$$(12) \qquad \hat{V}(\hat{\xi}) = (\hat{\Omega}_o^{-1} \otimes I) \sum_{t=2}^{T} (\hat{\tilde{u}}_t \hat{\tilde{u}}_t') \otimes (\hat{\tilde{u}}_{t-1} \hat{\tilde{u}}_{t-1}') (\hat{\Omega}_o^{-1} \otimes I)$$

Therefore the chi-square statistic associated with the score may be chosen as :

$$(13) \qquad S_1 = \left(\sum_{t=2}^{T} \hat{\tilde{u}}_t \otimes \hat{\tilde{u}}_{t-1}\right)' \left(\sum_{t=2}^{T} (\hat{\tilde{u}}_t \hat{\tilde{u}}_t') \otimes (\hat{\tilde{u}}_{t-1} \hat{\tilde{u}}_{t-1}')\right)^{-1} \left(\sum_{t=2}^{T} \hat{\tilde{u}}_t \otimes \hat{\tilde{u}}_{t-1}\right)$$

This statistic is distributed under the null, as a χ^2 with p^2 degrees of freedom. The score test procedure of the non correlation hypothesis Ho : (R=o) is :

$$(14) \qquad \begin{array}{ll} \text{accept Ho , if } S_1 < \chi^2_{1-\alpha}(p^2) \\ \text{reject Ho , if } S_1 > \chi^2_{1-\alpha}(p^2) \end{array}$$

4.e. Comparison with the D.W. Statistic.

In the unidimensional case, S_1 reduces to :

$$S_1 = \frac{\left(\sum_{t=2}^{T} \hat{\tilde{u}}_t \hat{\tilde{u}}_{t-1}\right)^2}{\sum_{t=2}^{T} \hat{\tilde{u}}_t^2 \hat{\tilde{u}}_{t-1}^2}$$

This form may be compared with the square of the autocorrelation coefficient of generalized residuals, i.e. with :

$$S_2 = T \frac{\left(\sum_{t=2}^{T} \hat{\tilde{u}}_t \hat{\tilde{u}}_{t-1}\right)^2}{\sum_{t=2}^{T} \hat{\tilde{u}}_t^2 \sum_{t=2}^{T} \hat{\tilde{u}}_{t-1}^2}$$

These two statistics are asymptotically equivalent under the null if and only if :

$$\text{plim} \frac{1}{T} \sum_{t=2}^{T} \hat{\tilde{u}}_t^2 \hat{\tilde{u}}_{t-1}^2 = \text{plim} \frac{1}{T} \sum_{t=2}^{T} \hat{\tilde{u}}_t^2 \, . \quad \text{plim} \frac{1}{T} \sum_{t=2}^{T} \hat{\tilde{u}}_{t-1}^2$$

$$\Longleftrightarrow \quad \text{plim} \frac{1}{T} \sum_{t=2}^{T} \tilde{u}_t^2 \tilde{u}_{t-1}^2 = \text{plim} \frac{1}{T} \sum_{t=2}^{T} \tilde{u}_t^2 \, . \quad \text{plim} \frac{1}{T} \sum_{t=2}^{T} \tilde{u}_{t-1}^2$$

This condition is not satisfied in general . However it is satisfied in two important cases :

i) if $y_t^* = y_t$, i.e if $\tilde{u}_t = u_t$

ii) if the successive observations of the exogenous variables x_t can be considered as i.i.d variables.

The first case covers the example of the usual linear model and this explains the usual relationship between the D.W. statistic and the score test statistic.

By analogy, it is clear that a generalized DURBIN-WASTON statistic has to be defined from S_1 and not from S_2, which is not in general equivalent to S_1. Such a statistic $\Delta.W.$ could be implicitely defined by :

$$S_1 = T\left(1 - \frac{\Delta.W.}{2}\right)^2$$

and would not in general be asymptotically equivalent to :

$$\frac{\sum_{t=2}^{T} (\hat{\tilde{u}}_t - \hat{\tilde{u}}_{t-1})^2}{\sum_{t=1}^{T} \hat{\tilde{u}}_t^2}$$

4.f. Score test with maintained nullity constraints on the autocorrelation matrix.

The previous test is easily generalized, when, in the maintained hypothesis, some elements of the autocorrelation matrix R are known to be equal to zero. The m unconstrained elements of R may be defined from $\text{vec}\, R$ through a selection matrix A, with size $m \times p^2$ and rank m ; each row of A has a unique non null element and this element is equal to 1.

In this case, the score test is based on the vector :

$$A \hat{\xi} = A \left[\frac{\partial \log \ell}{\partial \text{vec } R} (y_1 \ldots y_T; b, \Omega, R)\right]_{b = \hat{b}_{oT}, \Omega = \hat{\Omega}_{oT}, R = o}$$

The associated chi-square statistic has the following form :

$$(15) \qquad \boxed{S_1(A) = \hat{\xi}' A' [A \hat{V} \hat{\xi} A']^{-1} A \hat{\xi}}$$

where $\hat{\xi}$ and $\hat{V} \hat{\xi}$ are given by (11) and (12).

The statistic $S_1(A)$ depends in general on the variance-covariance matrix $\hat{\Omega}_o \otimes I$ appearing in $\hat{\xi}$. However, it is sometimes possible to simply suppress $\hat{\Omega}_o \otimes I$ in (15).

It is the case if and only if :

$$(\hat{\Omega}_o^{-1} \otimes I) A' [A(\hat{\Omega}_o^{-1} \otimes I) \hat{V} \hat{\xi} (\hat{\Omega}_o^{-1} \otimes I) A']^{-1} A(\hat{\Omega}_o^{-1} \otimes I) = A' [A \hat{V} \hat{\xi} A']^{-1} A$$

This condition is satisfied for any possible form of $\hat{V} \hat{\xi}$ iff the two matrices A' and $(\hat{\Omega}_o^{-1} \otimes I) A'$ have the same range :

$$\text{Rg } A' = \text{Rg } (\hat{\Omega}_o^{-1} \otimes I) A' \iff \text{Ker } A = \text{Ker } [A(\hat{\Omega}_o^{-1} \otimes I)]$$

$$\iff \quad (16) \qquad \boxed{\text{Ker } A = (\hat{\Omega}_o \otimes I) \text{ Ker } A}$$

As seen above, this necessary and sufficient condition is obviously satisfied when we are interested in all the m^2 elements of the autocorrelation matrix R, i.e when : $A = I$; it is also satisfied when the two matrices R and Ω have the same bloc diagonal representation, in particular when they are constrained to be diagonal. In such a case, the test statistic reduces to :

$$S_1(A) = \left(\sum_{t=2}^{T} \hat{\tilde{u}}_t \otimes \hat{\tilde{u}}_{t-1}\right)' A' \left\{A \sum_{t=2}^{T} (\hat{\tilde{u}}_t \hat{\tilde{u}}_t') \otimes (\hat{\tilde{u}}_{t-1} \hat{\tilde{u}}_{t-1}') A'\right\}^{-1} A\left(\sum_{t=2}^{T} \hat{\tilde{u}}_t \otimes \hat{\tilde{u}}_{t-1}\right)$$

If the selection matrix A can be decomposed into a tensorial product $A = B \otimes C$, this expression becomes :

$$S_1(B\otimes C) = \left(\sum_{t=2}^{T} \hat{\tilde{v}}_t \otimes \hat{\tilde{w}}_{t-1}\right)' \left\{\sum_{t=2}^{T} (\hat{\tilde{v}}_t \hat{\tilde{v}}_t') \otimes (\hat{\tilde{w}}_{t-1} \hat{\tilde{w}}_{t-1}')\right\}^{-1} \left(\sum_{t=2}^{T} \hat{\tilde{v}}_t \otimes \hat{\tilde{w}}_{t-1}\right)$$

with $\hat{\tilde{v}}_t = B \hat{\tilde{u}}_t$ and $\hat{\tilde{w}}_t = C \hat{\tilde{u}}_t$.

In particular if A is the $(1,p^2)$ matrix selecting the j^{th} diagonal term of R, we have

$$A = B_j \otimes B_j$$

where B_j is the p-dimensional row vector whose components are zero except the j^{th} component which is 1 . In this case :

$$S_1(B_j \otimes B_j) = \frac{\left(\sum_{t=2}^{T} \hat{\tilde{u}}_t^j \, \hat{\tilde{u}}_{t-1}^j\right)^2}{\sum_{t=2}^{T} (\hat{\tilde{u}}_t^j)^2 \, (\hat{\tilde{u}}_{t-1}^j)^2}$$

where $\hat{\tilde{u}}_t^j$ is the j^{th} component of $\hat{\tilde{u}}_t$.

Remark : The form of the statistic and its asymptotic properties remain the same if the parameter b and Ω are equality constrained. Thus it is in particular possible to fix equal to 1 some diagonal elements of Ω (case of qualitative models) or to impose to this matrix to be diagonal.

4.g. Determination of the predicted residuals.

The expression of the test statistic contains the predicted residuals $\hat{\tilde{u}}_t$. These residuals can be obtained along two different lines.

i) A first approach consists in computing the analytical form of the prediction of the disturbance and then to replace the parameters by their constrained M L estimations.

For instance, let us consider the disequilibrium model defined by :

$$\begin{cases} y_{1t}^{*} = x_{1t}\, b_1 + u_{1t} \\ y_{2t}^{*} = x_{2t}\, b_2 + u_{2t} \end{cases} \qquad \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \sim N\left|\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\right|$$

with $y_t = \text{Min}\,(y_{1t}^{*}, y_{2t}^{*})$. The prediction of u_{1t} in absence of autocorrelation is :

$$\begin{aligned} \tilde{u}_{1t} &= E_o(u_{1t}/y_t) \\ &= P_o[y_{1t}^{*} > y_{2t}^{*}/y_t] \quad E_o[u_{1t}/y_t, y_{1t}^{*} > y_{2t}^{*}] \\ &+ P_o[y_{2t}^{*} > y_{1t}^{*}/y_t] \quad E_o[u_{1t}/y_t, y_{2t}^{*} > y_{1t}^{*}] \\ &= P_o[y_{1t}^{*} > y_{2t}^{*}/y_t] \quad E_o[u_{1t}/y_t, y_{1t}^{*} > y_{2t}^{*}] \\ &+ P_o[y_{2t}^{*} > y_{1t}^{*}/y_t] \; (y_t - x_{1t}\, b_1) \end{aligned}$$

and the terms of this decomposition are :

$$P_o[y_{1t}^{*} > y_{2t}^{*}/y_t] = \frac{\Pi_{12}}{\Pi_{21} + \Pi_{12}}$$

with :
$$\Pi_{12} = \frac{1}{\sigma_2}\, \varphi\left(\frac{y_t - x_{2t}b_2}{\sigma_2}\right) \Phi\left(\frac{\frac{\sigma_{12}}{\sigma_2^2}(y_t - x_{2t}b_2) - y_t + x_{1t}b_1}{\sigma_1\sqrt{1-\rho^2}}\right)$$

$$\Pi_{21} = \frac{1}{\sigma_1}\, \varphi\left(\frac{y_t - x_{1t}b_1}{\sigma_1}\right) \Phi\left(\frac{\frac{\sigma_{12}}{\sigma_1^2}(y_t - x_{1t}b_1) - y_t + x_{2t}b_2}{\sigma_2\sqrt{1-\rho^2}}\right)$$

and :

$$E_o(u_{1t}/y_t, y_{1t}^{*} > y_{2t}^{*})$$

$$= \frac{\sigma_{12}}{\sigma_2^2}(y_t - x_{2t}b_2) + \sigma_1\sqrt{1-\rho^2}\; \lambda\left(\frac{y_t - x_{1t}b_1 - \frac{\sigma_{12}}{\sigma_2^2}(y_t - x_{2t}b_2)}{\sigma_1\sqrt{1-\rho^2}}\right)$$

where ρ is the correlation between u_{1t} and u_{2t}, φ and Φ are the density and the cumulative function of the standard normal and $\lambda = \frac{\varphi}{1 - \Phi}$ is the associated MILL's ratio.

This analytical approach has been used by ARTUS-LAROQUE MICHEL (1984) in their estimation of a French multi-markets disequilibrium model.

ii) Another approach appears to be easier, when the latent model is linear :

$$y_t^* = x_t b + u_t$$

In effect in such a case the constrained score :

$$\begin{bmatrix} \frac{\partial \log \ell(y;b,\Omega,o)}{\partial b} \\ \\ \frac{\partial \log \ell(y;b,\Omega,o)}{\partial \text{ vec } \Omega} \end{bmatrix}$$

, which is used to determine the constrained M.L estimators $\hat{b}_{oT}$ and $\hat{\Omega}_{oT}$, has a first component given by :

$$\frac{\partial \log \ell}{\partial b}(y;b,\Omega,o) = \sum_{t=1}^{T} \frac{\partial}{\partial b} \log \tilde{\ell}(y_t/x_t;b,\Omega,o) = \sum_{t=1}^{T} x_t' \tilde{u}_t$$

Therefore the generalized residuals $\hat{\tilde{u}}_t$ may in general be extracted from $\frac{\partial \log \ell}{\partial b}(y;\hat{b}_{oT},\hat{\Omega}_{oT},o)$ provided that each term of the sum $\sum_{t=1}^{T} \frac{\partial}{\partial b} \log \tilde{\ell}(y_t/x_t;b,\Omega,o)$ is computed.

5. ESTIMATION OF THE PARAMETERS.

As noted in section 3, it is in general impossible to compute the unconstrained M.L estimators of b,Ω,R . We propose in this section some consistent estimators of these parameters in presence of autocorrelation.

5.a. Constrained M.L. estimation of b and Ω .

The result obtained by ROBINSON (1982) in the context of tobit models can be generalized.

In effect, if autocorrelation is present, the constrained M L estimators $\hat{b}_{oT}$ and $\hat{\Omega}_{oT}$ respectively tends to the true value b_o of b and the true value of the variance of u_t :

$$Q_o = \sum_{i=o}^{\infty} R_o^i \, \Omega_o \, R_o^{i'}$$

(in particular $\hat{\Omega}_{oT}$ is not a consistent estimator of Ω_o).

Let us now see why this result holds.

Under the null hypothesis Ho : $(R=o)$, the log likelihood function is :

$$\log \ell(y_1 \ldots y_T; b,\Omega,o) = \sum_{t=1}^{T} \log \tilde{\ell}(y_t/x_t; b,\Omega)$$

where $\tilde{\ell}(y_t/x_t;b,\Omega)$ is the conditional density function of y_t given x_t under the null. The constrained M L estimators of b and Ω are solutions of :

$$\underset{b,\Omega}{\text{Max}} \ \frac{1}{T} \sum_{t=1}^{T} \log \tilde{\ell}(y_t/x_t;b,\Omega)$$

They converge to the solutions of the limit optimization problem (see BURGUETE-GALLANT-SOUZA (1982)).

$$\underset{b,\Omega}{\text{Max}} \ \text{plim} \ \frac{1}{T} \sum_{t=1}^{T} \overset{\circ}{E} \log \tilde{\ell}(y_t/x_t;b,\Omega)$$

where $\overset{\circ}{E}$ is the expectation with respect to the true conditional distribution of y_t given x_t, i.e the distribution associated with the true values b_o, Ω_o, R_o of the parameters.

We have :

$$\overset{\circ}{E} \log \tilde{\ell}(y_t/x_t;b,\Omega) = \int_y \log \tilde{\ell}(y_t/x_t;b,\Omega) \ \tilde{\ell}(y_t/x_t;b_o,Q_o) \ d\nu(y_t)$$

and, from KULLBACK's inequality, this quantity is maximum for $b = b_o$ and $\Omega = Q_o$, which gives the result.

Therefore the constrained M L estimator of b remains consistent when $R \neq o$, but the same result is not valid for the constrained M L estimator of the variance. More precisely $\hat{\Omega}_{oT}$ asymptotically overestimates Ω_o , since :

$$Q_o = \Omega_o + \sum_{i=1}^{\infty} R^i \Omega_o R^{i'} \gg \Omega_o$$

5.b. <u>Estimation of the autocorrelation.</u>

By analogy with the linear case, it might seem natural to consider the empirical correlation between the generalized residuals, i.e, in the unidimensional case :

$\tilde{\rho}_T = \dfrac{\sum_{t=2}^{T} \hat{\tilde{u}}_t \hat{\tilde{u}}_{t-1}}{\sum_{t=2}^{T} \hat{\tilde{u}}_T^2}$. However this quantity converges to the theoretical correlation between $\tilde{u}_t$ and $\tilde{u}_{t-1}$ and is in general different from the theoretical correlation ρ between the disturbances.

Another idea consists in directly studying the cross moment between the observables : $E(y_t \, y'_{t-1})$.

The cross moment $E(y_t \, y'_{t-1}) = E[g(y^*_t)g'(y^*_{t-1})]$ is a function of b_o, Q_o and R_o (since $\Omega_o = Q_o - R_o Q_o R'_o$)
$E(y_t \, y'_{t-1}) = \Psi(x_t, x_{t-1}, b_o, Q_o, R_o)$.

Since consistent estimates of b_o , Q_o have been found in the previous subsection, it is possible to consistently estimate the autocorrelation matrix R_o by mean of a quasi-generalized least squares method. Such an estimator $\tilde{R}_T$ could for instance be defined as a solution of :

$$(17) \qquad \underset{R}{\text{Min}} \sum_{t=2}^{T} \| \text{vec}\,(y_t\, y'_{t-1}) - \text{vec}\, \Psi(x_t, x_{t-1}, \hat{b}_{oT}, \hat{Q}_{oT}, R) \|^2 ,$$

if the elements of R are identifiable from the cross moment $E(y_t\, y'_{t-1})$.

Finally a consistent estimator of Ω_o may be obtained from the consistent estimators $\hat{\Omega}_{oT}$ and $\tilde{R}_T$ of Q_o and R_o .

<u>Example 18</u> : Let us consider a probit model with autocorrelation :

$$\begin{cases} y^*_t = x_t\, b + u_t \quad , \quad u_t = \rho\, u_{t-1} + \varepsilon_t \\ \varepsilon_t \sim N(o,1) \\ y_t = 1\!\!\mathrm{I}_{\, y^*_t > o} \end{cases}$$

The cross moment is given by :

$$E(y_t\, y'_{t-1}) = P(y^*_t > o\ ;\ y^*_{t-1} > o) = a(-\, x_t b_o,\ -\, x_{t-1} b_o, \rho_o)$$

where $a(\alpha, \beta, \rho) = P(v > \alpha,\ w > \beta)$

with $\begin{pmatrix} v \\ w \end{pmatrix} \sim N\left[\begin{pmatrix} o \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right]$.

Since a is strictly increasing with respect to ρ (see SLEPIAN (1962)), the solution of :

$$\underset{\rho}{\text{Min}} \sum_{t=2}^{T} (y_t \, y_{t-1} - a(- x_t \, \hat{b}_{oT}, - x_{t-1} \, \hat{b}_{oT}, \rho))^2$$

is unambiguously defined. This estimation method generalizes the tetrachoric estimation method introduced by PEARSON (1901) (see GOURIEROUX - MONFORT-TROGNON (1984)).

5.c. Consistency of the score test.

The knowledge of the asymptotic properties of $\hat{b}_{oT}$ and $\hat{\Omega}_{oT}$ under the alternative $(R_o \neq o)$ allows to study the consistency of the autocorrelation score test. Under usual regularity assumptions, a necessary and sufficient condition for the test to be consistent is :

$$\text{plim} \frac{1}{T} \hat{\xi}_T \neq o \qquad \text{under the alternative}$$

$$\iff \text{plim} \frac{1}{T} \sum_{t=2}^{T} \hat{\tilde{u}}_t \otimes \hat{\tilde{u}}_{t-1} \neq o$$

$$\iff \text{plim} \frac{1}{T} \sum_{t=2}^{T} \underset{\hat{b}_{oT}, \hat{\Omega}_{oT}, o}{E} (u_t / y_t) \otimes \underset{\hat{b}_{oT}, \hat{\Omega}_{oT}, o}{E} (u_{t-1} / y_{t-1}) \neq o$$

where $\underset{b,\Omega,R}{E}$ denotes an expectation with respect to the distribution associated with the values b,Ω,R of the parameters.

The previous condition is also equivalent to :

$$\text{plim}\ \frac{1}{T} \sum_{t=2}^{T} \underset{b_o,Q_o,o}{E} (u_t/y_t) \otimes \underset{b_o,Q_o,o}{E} (u_{t-1}/y_{t-1}) \neq o$$

$$\Longleftrightarrow \quad \text{plim}\ \frac{1}{T} \sum_{t=2}^{T} \underset{b_o,\Omega_o,R_o}{E} [\underset{b_o,Q_o,o}{E}(u_t/y_t) \otimes \underset{b_o,Q_o,o}{E}(u_{t-1}/y_{t-1})] \neq o$$

Noting that : $\underset{b_o,\Omega_o,R_o}{E} \ \underset{b_o,Q_o,o}{E}(u_t/y_t) = \underset{b_o,Q_o,o}{E} u_t = o$,

and that : $\underset{b_o,Q_o,o}{E} (u_t/y_t) = \underset{b_o,\Omega_o,R_o}{E} (u_t/y_t)$, we see that

the test is consistent, for any limit distribution of the exogenous variables, if and only if the predictions under the alternative $\underset{b_o,\Omega_o,R_o}{E} (u_t/y_t)$ and $\underset{b_o,\Omega_o,R_o}{E} (u_{t-1}/y_{t-1})$

are correlated.

Example 19 : In the probit model studied in example 18, the prediction is given by :

$$\frac{\varphi(x_t b_o)}{\Phi(x_t b_o)\ [1 - \Phi(x_t b_o)]} \quad (y_t - \Phi(x_t b_o))$$

Therefore, the covariance between successive predictions is proportional to :

$$\text{Cov}(y_t, y_{t-1}) = a(- x_t b_o, - x_{t-1} b_o, \rho_o) - \Phi(x_t b_o)\ \Phi(x_{t-1} b_o)$$

We deduce from the strict monotonicity of a with respect to ρ, that this quantity is equal to zero if and only if $\rho_0 = o$. This shows the consistency of the test.

6. SOME OTHER FORMULATIONS OF SERIAL CORRELATION.

In the previous sections, we modelize serial correlation by mean of an autoregressive scheme on the disturbances. We now consider some other ways of introducing such a correlation, restricting ourselves to the unidimensional case, for sake of simplicity.

6.a. Lagged endogenous variables.

The latent model is defined by :

$$y_t^* = \varphi \, y_{t-1}^* + f(x_t, b) + u_t$$

where the disturbances are i.i.d with normal distribution : $u_t \sim N(o, \sigma^2)$ and $|\varphi| < 1$; the observables are : $y_t = g(y_t^*)$.

The score statistic for testing the hypothesis of no serial correlation Ho : $(\varphi = o)$ is given by :

$$\hat{\xi}_T^{(1)} = \left[\frac{\partial \log\ell(y;b,\varphi,\sigma^2)}{\partial\varphi}\right]_{b_o=\hat{b}_{oT},\varphi=o,\sigma^2=\hat{\sigma}^2_{oT}}$$

$$= \left\{E\left[\frac{\partial \log\ell^*(y^*;b,\varphi,\sigma^2)}{\partial\varphi} \;/\; y\right]\right\}_{b=\hat{b}_{oT},\varphi=o,\sigma^2=\hat{\sigma}^2_{oT}}$$

$$= \left\{E\left[\frac{1}{\sigma^2}\sum_{t=2}^{T} u_t\, y^*_{t-1} \;/\; y\right]\right\}_{b=\hat{b}_{oT},\varphi=o,\sigma^2=\hat{\sigma}^2_{oT}}$$

$$= \frac{1}{\hat{\sigma}^2_{oT}} \sum_{t=2}^{T} \hat{\tilde{u}}_t \;(\hat{\tilde{u}}_{t-1} + f(x_{t-1},\hat{b}_{oT}))$$

For the same reasons as in 4.c. the statistic $\hat{\xi}_T^{(1)}$ is asymptotically equivalent, under Ho , to

$$\frac{1}{\sigma_o^2} \sum_{t=2}^{T} \tilde{u}_t \;(\tilde{u}_{t-1} + f(x_{t-1},b_o))$$

where $\tilde{u}_t = E_{\Theta_o}(u_t/y_t)$ with $\Theta_o = (b_o,o,\sigma_o^2)$

Since $E_{\Theta_o} \tilde{u}_t(\tilde{u}_{t-1} + f(x_{t-1},b_o)) = o$, the statistic $\frac{\sigma_o^2}{\sqrt{T}} \hat{\xi}_T^{(1)}$

is asymptotically zero mean normally distributed ; its asymptotic variance is $\sum_{j=-\infty}^{\infty} \Gamma_j$ with, for $j > o$

$$\Gamma_j = \operatorname{plim} \frac{1}{T} \sum_{t=2}^{T-j} E_{\Theta_o}[\tilde{u}_t\tilde{u}_{t+j}(\tilde{u}_{t-1}+f(x_{t-1},b_o)) \quad (\tilde{u}_{t-1+j}+f(x_{t-1+j}b_o))]$$

and $\Gamma_{-j} = \Gamma_j$.

It is clear that $\Gamma_j = o \ \forall \ j \neq o$ and, therefore, the asymptotic variance of $\frac{\sigma_o^2}{\sqrt{T}} \ \hat{\xi}_T^{(1)}$ is

$$\text{plim} \ \frac{1}{T} \sum_{t=2}^{T} E_{\Theta_o} \tilde{u}_t^2 \ (\tilde{u}_{t-1} + f(x_{t-1}, b_o))^2$$

$$= \ \text{plim} \ \frac{1}{T} \sum_{t=2}^{T} \tilde{u}_t^2 \ \tilde{y}_{t-1}^2$$

with $\tilde{y}_t \ = \ E_{\Theta_o}(y_t^* / y_t)$

Finally the statistic $\hat{\xi}_T^{(1)}$ is such that, under Ho :

$$\frac{1}{T} \ \hat{\xi}_T^{(1)} \xrightarrow[T \to \infty]{a.s} \ o \quad \text{and} \quad \frac{1}{\sqrt{T}} \ \hat{\xi}_T^{(1)} \xrightarrow[T \to \infty]{d} \ N \left[o, \ \frac{1}{\sigma_o^4} \ \text{plim} \ \frac{1}{T} \sum_{t=2}^{T} \hat{\tilde{u}}_t^2 \ \tilde{y}_{t-1}^2 \right].$$

The score statistic is :

$$\frac{\left(\sum_{t=2}^{T} \hat{\tilde{u}}_t \ \hat{\tilde{y}}_{t-1} \right)^2}{\sum_{t=2}^{T} \hat{\tilde{u}}_t^2 \ \hat{\tilde{y}}_{t-1}^2}$$

whose asymptotic distribution, under Ho, is $\chi^2(1)$ and the test follows. In the same particular cases as in 4.e , $\sum_{t=2}^{T} \hat{\tilde{u}}_t^2 \ \hat{\tilde{y}}_{t-1}^2$ may be replaced by $\frac{1}{T} \left(\sum_{t=2}^{T} \hat{\tilde{u}}_t^2 \right) \left(\sum_{t=2}^{T} \hat{\tilde{y}}_{t-1}^2 \right)$.

Let us now briefly consider the estimation problem when Ho is uncorrectly assumed. Under the null, the log likelihood function is $\log \ell(y_1, \ldots, y_T; b, o, \sigma^2) \ = \ \sum_{t=1}^{T} \log \tilde{\ell}(y_t / x_t; b, \sigma^2)$ where $\tilde{\ell}(y_t / x_t; b, \sigma^2)$ is the conditional density function of y_t given x_t , under H_o .

The M L estimators of b and σ^2 under Ho are solutions of

$$\underset{b,\sigma^2}{\text{Max}} \frac{1}{T} \sum_{t=1}^{T} \log \tilde{\ell}(y_t/x_t;b,\sigma^2)$$

and these estimators converge to the solutions of

$$\underset{b,\sigma^2}{\text{Max}} \quad \text{plim} \frac{1}{T} \sum_{t=1}^{T} \overset{\circ}{E} \log \tilde{\ell}(y_t/x_t;b,\sigma^2)$$

where $\overset{\circ}{E}$ is the expectation with respect to the true conditional distribution of y_t given x_t. The density function of this distribution is not, in general, of the same form as $\tilde{\ell}$ and the argument of 5.a does not hold. Note however that if the x_t's are i.i.d normally distributed and if $f(x_t,b) = x_t b = \beta + \sum_{i=1}^{K} x_{it} \beta_i$, the conditional distribution of y_t^* given x_t is :

$$N\left(x_t b + \frac{\varphi}{1 - \varphi} E\, x.b \quad , \quad \frac{\sigma^2 + \varphi^2\, b'\, Vx\, b}{1 - \varphi^2}\right)$$

the same argument as in 5.a, shows that the estimator of $(\beta_1,\ldots,\beta_K)$ under $\varphi = o$ remains consistent when $\varphi \neq o$, but it is not the case for β and σ^2.

B.b. General autoregressive disturbances.

Let us now consider a latent model defined by

$$h(y_t^*,x_t;b) = u_t \qquad t = 1,\ldots,T$$

where u_t is a zero-mean, gaussian autoregressive process of order q ; this process is defined by :

$$u_t - \varphi_1 u_{t-1} - \ldots - \varphi_q u_{t-q} = \varepsilon_t$$

or, using the lag operator B , by :

$$\Phi(B) u_t = \varepsilon_t$$

with $\Phi(B) = 1 - \varphi_1 B - \ldots - \varphi_q B^q$

where the ε_t's are $IIN(o,\sigma^2)$ and the roots of $\Phi(B)$ are assumed to be outside the unit circle. The observables are $y_t = g(y_t^*)$ and Ho is defined by $\varphi' = (\varphi_1,\ldots,\varphi_q) = o$.

The likelihood function of the latent model is :

$$\ell^*(y^*;b,\varphi,\sigma^2) = (2\pi)^{-\frac{T}{2}} (\det W)^{-\frac{1}{2}} \prod_{t=1}^{T} J_t(y_t^*,b)\exp(-\tfrac{1}{2} u' W^{-1} u)$$

where $J_t(y_t^*,b) = \left|\det \frac{\partial h(y_t^*,x_t;b)}{\partial y_t^{*\prime}}\right|$

$$u' = (u_1,\ldots,u_T)$$

$$W = V(u) = \sigma_u^2 \begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{T-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{T-2} \\ & & & & \\ & & & & \rho_1 \\ \rho_{T-1} & \cdots & \cdots & \rho_1 & 1 \end{pmatrix}$$

σ_u^2 is the variance of u_t and ρ_h is the correlation coefficient between u_t and u_{t-h} .

The score vector with respect to φ in the latent model is such that :

$$\frac{\partial \ \mathrm{Log}\ell^*}{\partial \varphi_i} = -\frac{1}{2}\ \frac{\partial}{\partial \varphi_i}\ \log\ (\det W) - \frac{1}{2}\ \frac{\partial}{\partial \varphi_i}\ u'\ W^{-1}\ u$$

The first term is equal to zero under Ho ; in effect,

$$\left(\frac{\partial}{\partial \varphi_i}\ \mathrm{Log}\ \det W\right)_{Ho} = \mathrm{tr}\ \left(W^{-1}\ \frac{\partial W}{\partial \varphi_i}\right)_{Ho}$$

$$= \frac{1}{\sigma^2}\ \mathrm{tr}\ \left(\frac{\partial W}{\partial \varphi_i}\right)_{Ho} = \frac{T}{\sigma^2}\ \left(\frac{\partial \sigma_u^2}{\partial \varphi_i}\right)_{Ho}$$

moreover :

$$\frac{\partial \sigma_u^2}{\partial \varphi_i} = \frac{\partial [\sigma^2/(1-\varphi_1\rho_1 - \ldots - \varphi_q\rho_q)]}{\partial \varphi_i}$$

$$= \frac{\sigma^2 \rho_i}{(1-\varphi_1\rho_1 - \ldots - \varphi_q\rho_q)^2} + \sigma^2 \sum_{j=1}^{q} \left(\frac{\partial \rho_j}{\partial \varphi_i}\right) \frac{\varphi_j}{(1-\varphi_1\rho_1 - \ldots - \varphi_q\rho_q)^2}$$

and : $\left(\frac{\partial \sigma_u^2}{\partial \varphi_i}\right)_{Ho} = o$ since $\rho_i = o \ \forall\ i > o$ under Ho .

Let us now consider the term $\left(\frac{\partial}{\partial \varphi_i}\ u'\ W^{-1}\ u\right)_{Ho}$

$$\frac{\partial}{\partial \varphi_i} (u' \, W^{-1} \, u) = \frac{\partial}{\partial \varphi_i} (tr \, W^{-1} \, u \, u')$$

$$= tr \left(\frac{\partial W^{-1}}{\partial \varphi_i} \; . \; u \, u' \right)$$

We have :

$$\frac{\partial W^{-1}}{\partial \varphi_i} = - W^{-1} \frac{\partial W}{\partial \varphi_i} W^{-1}$$

$$\left(\frac{\partial W^{-1}}{\partial \varphi_i} \right)_{Ho} = - \frac{1}{\sigma^4} \left(\frac{\partial W}{\partial \varphi_i} \right)_{Ho}$$

Using the result $\left(\frac{\partial \sigma_u^2}{\partial \varphi_i} \right)_{Ho} = 0$ we get :

$$\left(\frac{\partial W}{\partial \varphi_i} \right)_{Ho} = \sigma^2 \left[\frac{\partial}{\partial \varphi_i} \begin{pmatrix} 1 & \rho_1 & \cdots\cdots & \rho_{T-1} \\ & & & \\ \rho_{T-1} & \cdots & \cdots & 1 \end{pmatrix} \right]_{Ho}$$

Moreover we have :

$$\begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_T \end{pmatrix} = \begin{pmatrix} 1 & \rho_1 & \cdots\cdots & \rho_{q-1} \\ \rho_1 & 1 & & \rho_{q-2} \\ & & & \\ \rho_{T-1} & \rho_{T-2} & \cdots\cdots & \rho_{T-q} \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \vdots \\ \vdots \\ \varphi_q \end{pmatrix}$$

differentiating with respect to φ_i and evaluating the result under Ho we obtain :

$$\begin{pmatrix} \frac{\partial \rho_1}{\partial \varphi_i} \\ \vdots \\ \vdots \\ \frac{\partial \rho_T}{\partial \varphi_i} \end{pmatrix}_{Ho} = \left[\frac{\partial}{\partial \varphi_i} \begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{q-1} \\ & & & \\ & & & \\ \rho_{T-1} & \cdots & \cdots & \rho_{T-q} \end{pmatrix} \right]_{Ho} \begin{pmatrix} 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \\ \hline 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$$

which shows that

$$\left(\frac{\partial \rho_h}{\partial \varphi_i}\right)_{Ho} = \begin{cases} 1 & \text{if } h = i \\ 0 & \text{otherwise} \end{cases}$$

Finally we have :

$$\left(\frac{\partial}{\partial \varphi_i} u' W^{-1} u\right)_{Ho} = - \frac{1}{\sigma^2} \operatorname{tr} (A_i u u')$$

$$= - \frac{1}{\sigma^2} u' A_i u$$

where A_i is the $(T \times T)$ matrix whose terms are

$$a_{hk} = \begin{cases} 1 & \text{if } |h - k| = i \\ 0 & \text{otherwise} \end{cases}$$

This implies :

$$\left(\frac{\partial}{\partial \varphi_i} u' W^{-1} u\right)_{Ho} = - \frac{2}{\sigma^2} \sum_{t=i+1}^{T} u_t u_{t-i}$$

and :

$$\left(\frac{\partial \log \ell^*}{\partial \varphi_i}\right)_{Ho} = \frac{1}{\sigma^2} \sum_{t=i+1}^{T} u_t \, u_{t-i}$$

From (8) we get the components of the score vector :

$$\left(\frac{\partial \operatorname{Log} \ell}{\partial \varphi_i}\right)_{Ho} = \frac{1}{\sigma^2} \sum_{t=i+1}^{T} E_{\Theta_o}(u_t \, u_{t-i}/y)$$

$$= \frac{1}{\sigma^2} \sum_{t=i+1}^{T} E_{\Theta_o}(u_t/y_t) \, E_{\Theta_o}(u_{t-i}/y_{t-i})$$

The score statistic for testing Ho is the q-dimensional vector $\hat{\xi}_T^{(2)}$ whose components are :

$$\frac{1}{\hat{\sigma}^2_{oT}} \sum_{t=i+1}^{T} \hat{\tilde{u}}_t \, \hat{\tilde{u}}_{t-i} \qquad i = 1,\dots,q$$

The same argument as in 4.c shows that $\hat{\xi}_T^{(2)}$ is asymptotically equivalent, under Ho , to the vector whose components are :

$$\frac{1}{\sigma_o^2} \sum_{t=q+1}^{T} \tilde{u}_t \, \tilde{u}_{t-i}$$

with $\tilde{u}_t = E_{\Theta_o}(u_t/y_t)$

Since $E_{\Theta_o} \tilde{u}_t \, \tilde{u}_{t-i} = o$, $\frac{\sigma_o^2}{\sqrt{T}} \hat{\xi}_T^{(2)}$ is asymptotically zero mean normally distributed ; the asymptotic covariance matrix is $\sum_{j=-\infty}^{\infty} \Gamma_j$ with, for $j > o$

$$\Gamma_j = \text{plim} \frac{1}{T} \sum_{t=q+1}^{T-j} E_{\Theta_o} (\tilde{u}_t \, \tilde{u}_{t-q}^{t-1} \, \tilde{u}_{t+j} \, \tilde{u}'^{\,t+j-1}_{t+j-q})$$

and $\Gamma_{-j} = \Gamma'_j$

where $\tilde{u}_k^{\ell}$ is the vector whose components are $\tilde{u}_k, \tilde{u}_{k+1}, \ldots, \tilde{u}_{\ell}$.

The typical element of Γ_j is

$$E_{\Theta_o} (\tilde{u}_t \, \tilde{u}_{t-k} \, \tilde{u}_{t+j} \, \tilde{u}_{t+j-\ell}) \qquad \begin{matrix} 1 < k < q \\ 1 < \ell < q \end{matrix}$$

and, therefore, $\Gamma_j = o \;\; \forall \; j \neq o$. It follows that the asymptotic covariance of $\frac{\sigma_o^2}{\sqrt{T}} \; \hat{\xi}_T^{(2)}$ is the diagonal matrix Γ_o whose diagonal elements are

$$\text{plim} \frac{1}{T} \sum_{t=q+1}^{T} E_{\Theta_o} \tilde{u}_t^2 \, \tilde{u}_{t-i}^2 = \text{plim} \frac{1}{T} \sum_{q+1}^{T} \tilde{u}_t^2 \, \tilde{u}_{t-i}^2 \qquad i = 1, \ldots, q \; .$$

Finally the test of Ho : $\varphi = o$ is based on the statistic :

$$\sum_{i=1}^{q} \frac{\left(\sum_{t=i+1}^{T} \hat{\tilde{u}}_t \, \hat{\tilde{u}}_{t-i} \right)^2}{\sum_{t=i+1}^{T} \hat{\tilde{u}}_t^2 \, \hat{\tilde{u}}_{t-i}^2}$$

whose distribution under Ho is $\chi^2(q)$ and the test procedure follows. Note that, except in the special cases discussed in 4.e., the i^{th} term of the previous sum is not equal to T times the square of the correlation coefficient between the generalized residuals $\hat{\tilde{u}}_t$ and $\hat{\tilde{u}}_{t-i}$.

On the estimation side, the same argument as in 5.a shows that the maximum likelihood estimator $(\hat{b}_{oT}, \hat{\sigma}^2_{oT})$ of (b, σ^2), constrained by $\varphi = o$, is such that $\hat{b}_{oT}$ remains consistent if $\varphi \neq o$, whereas $\hat{\sigma}^2_{oT}$ converges to $\sigma^2_u \neq \sigma^2$.
More precisely

$$\sigma^2_u = \frac{\sigma^2}{1 - \rho_1\varphi_1 - \ldots - \rho_q\varphi_q} = \sigma^2 \left[1 + \sum_{i=1}^{\infty} \Psi_i^2\right] > \sigma^2$$

where the Ψ_i's are the coefficients of the long division of 1 by $\Phi(B)$; therefore, for T sufficiently large, $\hat{\sigma}^2_{oT}$ will overestimate σ^2.

6.c. <u>General moving average disturbances</u>

Let us now assume that the model is the same as in the previous subsection except that the u_t process is assumed to be a moving-average process of order q defined by :

$$u_t = \varepsilon_t + \gamma_1 \varepsilon_{t-1} + \ldots + \gamma_q \varepsilon_{t-q}$$

$$= \gamma(B) \varepsilon_t$$

It is easily seen that the same arguments holds provided that the matrix W is replaced by a matrix V whose elements are (with the convention $\gamma_o = o$) :

$$v_{ij} = \sigma^2 \sum_{k=|i-j|}^{q} \gamma_k \gamma_{k-|i-j|} \qquad \text{if } |i-j| < q$$

$$= 0 \qquad \text{otherwise}$$

A similar algebra shows that, under Ho : $\{\gamma_i = 0 \ , \ i = 1,\ldots,q\}$, the score vector associated with the parameters γ_i is identical to the score vector of the previous subsection ; in other words we have

$$\left(\frac{\partial \ \text{Log}\ell}{\partial \gamma_i}\right)_{Ho} = \frac{1}{\sigma^2} \sum_{t=i+1}^{T} E_{\Theta_0}(u_t/y_t)\ E_{\Theta_0}(u_{t-i}/y_{t-i})$$

It follows that exactly the same statistic can be used for testing H_0 against autoregressive or against moving average scheme (compare GODFREY (1978), GODFREY and WICKENS (1982)).

The remarks concerning the estimation remain also valid.

R E F E R E N C E S

ARTUS P., G. LAROQUE and G. MICHEL (1984)
Estimation of a quaterly macroeconomic model with quantity rationing.
to appear in Econometrica.

BURGUETE J., GALLANT R. and G. SOUZA (1982)
On unification of the asymptotic theory of non linear econometrics.
Econometric review.

DAGENAIS M. (1982)
The tobit model with serial correlation.
Economic Letters, 10 n° 3-4, 263-267.

DURBIN J. and G. WATSON (1951)
Testing for serial correlation in least squares regression.
Biometrika, vol. 38, 159-178.

GODFREY L.G. (1978)
Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables.
Econometrica 46, 1293-1302.

GODFREY L.G. and M.R. WICKENS (1982)
Tests of misspecification using locally equivalent alternative models.
In "Evaluating the reliability of macro economic models", Wiley.

GOURIEROUX C., LAFFONT J.J. and A. MONFORT (1980)
Coherency conditions in simultaneous linear equation models with endogenous switching regimes.
Econometrica 48, 675-695.

GOURIEROUX C., MONFORT A. and A. TROGNON (1984)
Estimation and test in probit models with serial correlation.
In "Alternative approaches to time series analysis" Publications des Facultés Universitaires Saint-Louis, Bruxelles.

MAC RAE KENNAN D. (1982)
A time series analysis of binary data.
Journal of the American Statistical Association, 77, 816-821.

PEARSON K. (1901)
On the correlation of characters not quantitatively measurable.
Philosophical Transactions of the Royal Society, 195, A.

ROBINSON P. (1982)
On the asymptotic properties of estimators of models containing limited dependent variables.
Econometrica 50, 27-41.

SLEPIAN D. (1962)
The one sided barrier problem for gaussian noise.
Bell System Technical Journal, 41, 463-501.

WHITE H. (1982)
Maximum likelihood of misspecified models.
Econometrica 50, 1-26.