N° 8129

PSEUDO MAXIMUM LIKELIHOOD METHODS : THEORY

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

* Universite PARIS IX-DAUPHINE et CEPREMAP, Paris.

** ECOLE NATIONALE DE LA STATISTIQUE ET DE L'ADMINISTRATION ECONOMIQUE, Malakoff.

1. INTRODUCTION

In this paper we are interested in the properties of the estimators obtained by maximizing a likelihood function associated with a family of probability distributions, which does not necesseraly contains the true distribution ; this kind of method will be called pseudo maximum likelihood method.

More precisely, we determine the families of pseudo likelihood function providing consistent and asymptotically normal estimators of parameters involved in the true distribution.

After some preliminaries in section 2, we define in section 3 the exponential families of type I and we show in section 4 that this kind of family gives consistent and asymptotically normal estimators of the parameters appearing in the first order moment of the true distribution ; moreover it is shown that, conversely, any family with this property is necessarily exponential of type I.

In section 5, we propose a generalization of the previous method adapted to the case in which it is possible to estimate consistently the second order moments of the true distribution. Finally, in section 6, we define the exponential families of type II and we show that these families are the only ones providing consistent estimators of the parameters appearing in the first and the second moments. Some technical proofs are gathered in four appendices and the application of the method of scoring is discussed in another one.

2. PREMILINARIES

The approach used in this paper rest upon previous results by JENNRICH (1969), MALINVAUD (1970), GALLANT-HOLLY (1980) and BURGUETE-GALLANT (1980) ; these results establish the asymptotic properties of estimators obtained by maximizing a function.

Let us consider G-variate vectors y_t , $t=1,\dots,T$ generated by the model :

$$(1) \qquad y_t = f(x_t,\theta_0) + e_t$$

where $\theta_0 \in \Theta \subset \mathbb{R}^k$, $x_t \in \mathbb{R}^m$, $y_t \in \mathbb{R}^G$, $e_t \in \mathbb{R}^G$.
We shall assume that the conditional distribution of $e_1,\dots,e_T$ given $x_1,\dots,x_T$ is equal to the product of the conditional distributions $L(e_t|x_t)$, where $L(e_t|x_t=x) = L(e_\tau|x_\tau=x)$ for $t \neq \tau$.
The true conditional distribution of $y_t|x_t$, which is unknown, will be denoted by $\lambda_0(x_t,\theta_0)$; it will be assumed that the expectation of $\lambda_0(x_t,\theta_0)$ is $f(x_t,\theta_0)$ and that the covariance matrix $\Sigma_0(x_t)$ exists for any x_t ; E_0 will denote the mathematical expectation with respect to this distribution. We want to extimate θ_0 by considering the solutions of the problem:

$$\underset{\theta\in\Theta}{\mathrm{Max}} \sum_{t=1}^{T} \varphi(y_t,x_t,\theta)$$

Under classical assumptions (denoted by $\mathcal{A}$) such as : f and φ are continuous with respect to all the variables and twice continuously differentiable with respect to θ , Θ is a compact set, θ_0 is in the interior of Θ , almost every realization of (e_t, x_t) generates Cesaro summable sequences... (see GALLANT-HOLLY and BURGUETE-GALLANT) it can be shown :

a) $\varphi_T(y^T, x^T, \theta) = \frac{1}{T} \sum_{t=1}^{T} \varphi(y_t, x_t, \theta)$ converges almost surely, uniformly on Θ , to $\varphi_\infty(\theta) = \underset{x}{E} \underset{o}{E} \varphi(y, x, \theta)$ where $\underset{x}{E}$ is the expectation with respect to a probability measure μ not depending on φ , θ and f .

b) $\varphi_\infty(\theta)$ is twice differentiable with respect to θ and the partial derivatives of φ_∞ may be obtained by inverting the expectation and derivation operators.

c) If φ_∞ has a unique maximum in θ_0 , then the estimators $\hat{\theta}_T$ obtained by maximizing $\varphi_T(y^T, x^T, \theta)$ exist almost surely and converge a.s. to θ_0 ; moreover $\sqrt{T}(\hat{\theta}_T - \theta_0)$ converges in distribution to $N(0, J^{-1} I J^{-1})$ where

$$J = \underset{x}{E} \underset{o}{E} \left[- \frac{\partial^2 \varphi(y, x, \theta_0)}{\partial\theta \, \partial\theta'} \right]$$

and $$I = \underset{x}{E} \underset{o}{E} \left[\frac{\partial \varphi}{\partial \theta}(y, x, \theta_0) \frac{\partial \varphi}{\partial \theta}(y, x, \theta_0)' \right]$$

These results can be extended to the case where φ also depends on a nuisance parameter η . If the "nuisance parameter" η is replaced by $\hat{\eta}_T$ converging a.s. to some value η_1 , then the estimators

$\hat{\theta}_T$ obtained by maximising $\frac{1}{T}\sum_{t=1}^{T}\varphi(y_t,x_t,\theta,\hat{\eta}_T)$ have exactly the same properties as before if η is replaced by η_1 is φ_∞ , J and I (See BURGUETE-GALLANT)

3. EXPONENTIAL FAMILIES OF TYPE I

In this section we shall use classical results on FISHER information, kullback information, identification and exponential family ; these results may be found in MONFORT (1981).

Définition 1 : A family of probability measures on $\mathbb{R}^G$, indexed by a parameter $m \in M \subset \mathbb{R}^G$ is called exponential of type I, if :

a) every element of the family has a density function with respect to a given measure $\nu(du)$; this density function may be written as :
$\ell(u,m) = \exp\{A(m) + B(u) + C(m)\,u\}$, $u \in \mathbb{R}^G$,
where $A(m)$ and $B(u)$ are scalar and $C(m)$ is a row vector of size G .

b) m is the mean of the distribution, whose density is equal to $\ell(u,m)$, and this parameter is identifiable.

Note that a) implies that all the moments exist.
We also assume that conditions α are satisfied for f such that $f(x,\theta) \in M\ \forall x , \forall \theta$ and for φ defined by : $\varphi(y,x,\theta) = \text{Log}\ \ell[y,f(x,\theta)]$
This implies in particular, that the functions A and C are twice continuously differentiable on $\overset{\circ}{M}$, the interior of M .

Property 1 : If $(\ell(u,m), m \in M)$ is an exponential family of type I, we have :

$$\frac{\partial A(m)}{\partial m} + \frac{\partial C(m)}{\partial m}\, m = 0$$

Proof : Differentiating the equality :

$$\int \ell(u,m)\ \nu(du) = 1$$, we obtain :

$$\int \left(\frac{\partial A(m)}{\partial m} + \frac{\partial C(m)}{\partial m} u \right) \ell(u,m)\ \nu(du) = 0$$

or $$\frac{\partial A(m)}{\partial m} + \frac{\partial C(m)}{\partial m} m = 0$$ Q.E.D.

Property 2 : $$\frac{\partial^2 A}{\partial m\, \partial m'} + \sum_{g=1}^{G} \frac{\partial^2 C_g}{\partial m\, \partial m'} m_g + \frac{\partial C(m)}{\partial m} = 0$$

where C_g and m_g are respectively the g^{th} component of C and m .

Proof : : This property is obtained by differentiating the equality of property 1. Q.E.D.

Property 3 : $\frac{\partial C}{\partial m} = \Sigma^{-1}$, where Σ is the covariance matrix associated with $\ell(u,m)$.

Proof : From part b) of definition 1, we have

$$\int u\ \ell(u,m)\ \nu(du) = m$$

Differentiating this equality, we obtain :

$$\int u \left(\frac{\partial A'}{\partial m} + u' \frac{\partial C'}{\partial m} \right) \ell(u,m)\nu(du) = I_G$$

where I_G is the identity matrix.

Using property 1, we deduce that :

$$\int u(u-m)' \frac{\partial C'}{\partial m} \ell(u,m)\nu(du) = I_G$$

$$\Longleftrightarrow \Sigma \frac{\partial C'}{\partial m} = I_G$$

$$\Longleftrightarrow \frac{\partial C}{\partial m} = \Sigma^{-1} .$$ Q.E.D.

Remark that Σ is invertible, since in exponential families the identifiability of m is equivalent to the non singularity of the information matrix.

Property 4 : $\forall\, m \;,\; m_0 \in \mathcal{M}$, we have :

$$A(m) + C(m)\, m_0 \;\le\; A(m_0) + C(m_0)\, m_0$$

The equality holds, if and only if $m = m_0$

Proof : Kullback's inequality implies :

$$\int \mathrm{Log}\, \ell(u,m)\, \ell(u,m_0)\nu(du) \;\le\; \int \mathrm{Log}\, \ell(u,m_0)\, \ell(u,m_0)\nu(du)$$

Or : $A(m) + C(m)\, m_0 \;\le\; A(m_0) + C(m_0)\, m_0$

The equality is possible if and only if : $\ell(u,m) = \ell(u,m_0)$ ν a.s.,

or, since m is identifiable, if and only if $m = m_0$. Q.E.D.

Many classical families of probability measures are exponential of type I. Some examples of such families are given in table 1.

TABLE 1 : EXAMPLE OF EXPONENTIAL FAMILY OF TYPE I

Family	M	density function	C(m)
Binomial (n given)	$]0,n[$	$\frac{\Gamma(n+1)}{\Gamma(u+1)\Gamma(n-u+1)}\left(\frac{m}{n}\right)^{u}\left(1-\frac{m}{n}\right)^{1-u}$	$\text{Log}\,\frac{m}{n-m}$
Poisson	$\mathbb{R}^{+*}$	$\frac{e^{-m}\,m^{u}}{u!}$	$\text{Log}\,m$
Negative binomial (α given)	$\mathbb{R}^{+*}$	$\frac{\Gamma(\alpha+u)}{\Gamma(\alpha)\Gamma(u+1)}\left(\frac{m}{\alpha}\right)^{u}\left(1+\frac{m}{\alpha}\right)^{-(\alpha+u)}$	$\text{Log}\,\frac{m}{\alpha+m}$
Gamma (α given)	$\mathbb{R}^{+*}$	$\frac{u^{\alpha-1}\,e^{-\frac{u}{m}\alpha}}{\Gamma(\alpha)\left(\frac{m}{\alpha}\right)^{\alpha}}$	$-\frac{\alpha}{m}$
Normal (σ given)	$\mathbb{R}$	$\frac{1}{\sigma\sqrt{2\Pi}}\exp\left[-\frac{1}{2}\frac{(y-m)^2}{\sigma^2}\right]$	$\frac{m}{\sigma^2}$
Multinomial (n given)	$\sum_g m_g=n$, $m_g>0$	$\frac{n!}{\prod_g (u_g)!}\prod_g\left(\frac{m_g}{n}\right)^{u_g}$	$\text{Log}\,\frac{m_1}{n},\ldots,\text{Log}\,\frac{m_g}{n}$
Normal multivariate (Ω given)	$\mathbb{R}^G$	$\frac{\exp-\frac{1}{2}(u-m)'\Omega^{-1}(u-m)}{(2\Pi)^{G/2}\sqrt{\det\Omega}}$	$m'\Omega^{-1}$

The multivariate generalizations of Poisson, Negative, Binomial distributions (JOHNSON-KOTZ, Chap. 11) are also exponential families of type I.

4. PSEUDO-MAXIMUM LIKELIHOOD ESTIMATOR OF TYPE I

Let us now turn to the problem of estimating θ_0 in model (1) : $y_t = f(x_t,\theta_0) + e_t$, defined in section 2, assuming furthermore the first order identification of θ : $f(x,\theta_1) = f(x,\theta_2)$ μ. as $\Longrightarrow \theta_1 = \theta_2$. We are interested in pseudo maximum likelihood estimator of θ_0, obtained by maximizing $\sum_{t=1}^{T} \text{Log}\, \ell[y_t,f(x_t,\theta)]$, where $\ell(u,m)$ is a family of p.d.f. indexed by the mean $m \in M$, satisfying assumptions $\mathcal{A}$ and where $f(x,\theta) \in M$, $\forall x \forall \theta$.

<u>THEOREM 1</u> : Under the previous assumptions and if $(\ell(u,m), m \in M)$ is an exponential family of type I, the pseudo-maximum likelihood estimator (P.M.L.E.) of θ_0 is strongly consistent.

<u>Proof</u> : Using the results mentioned in preliminaries, the strong consistency will be established, if it is shown, that limit function :

$$\varphi_\infty(\theta) = \underset{x}{E}\,\underset{o}{E}\, \text{Log}\, \ell[y,f(x,\theta)] \text{ has a unique maximum at } \theta_0 .$$

It is the case, since $\varphi_\infty(\theta)$ may be written :

$$\varphi_\infty(\theta) = \underset{x}{E}\, \{\underset{o}{E}\, B(y) + A[f(x,\theta)] + C[f(x,\theta)]\, f(x,\theta_0)\}$$

and the result directly follows from property 4 and from the first order identifiability of θ . Q.E.D.

Note that it is equivalent to maximize $\sum_{t=1}^{T} \text{Log}\, \ell(y_t,f(x_t,\theta))$ or to maximize $\sum_{t=1}^{T} \{A[f(x_t,\theta)] + C[f(x_t,\theta)]y_t\}$. Therefore it is not necessary to impose on y_t the constraints, which may be implied by the definition of B . For instance, the pseudo-maximum likelihood method with Poisson family may be applied even if y_t is any real number ; however, in that case, $f(x_t,\theta)$ must be positive, for any t .

The objective function which is maximized in the P.M.L. approach depends on the retained family, and it may be interesting to exhibit the objective functions associated with the classical families.

TABLE 2 : OBJECTIVE FUNCTIONS

Family	Objective function
Binomial (n given, $n \in \mathbb{N}$)	$\sum_{t=1}^{T} \{\text{Log}\left[1 - \frac{f(x_t,\theta)}{n}\right] + y_t \text{ Log}\left[\frac{f(x_t,\theta)}{n - f(x_t,\theta)}\right]\}$
Poisson	$\sum_{t=1}^{T} \{ - f(x_t,\theta) + y_t \text{ Log } f(x_t,\theta)\}$
Negative Binomial (α given)	$\sum_{t=1}^{T} \left[- \alpha \text{ Log}\left(1 + \frac{f(x_t,\theta)}{\alpha}\right) + y_t \text{ Log}\left(\frac{f(x_t,\theta)}{\alpha + f(x_t,\theta)}\right)\right]$
Gamma (α given)	$\sum_{t=1}^{T} \left(- \alpha \text{ Log } f(x_t,\theta) - \frac{\alpha\, y_t}{f(x_t,\theta)}\right)$
Normal (σ given)	$\sum_{t=1}^{T} [y_t - f(x_t,\theta)]^2$
Multinomial (n given, $n \in \mathbb{N}$)	$\sum_{t=1}^{T} \sum_{g=1}^{G} y_{gt} \text{ Log } f_g(x_t,\theta)$ where $\sum_{g} f_g(x_t,\theta) = n$
Normal multivariate (Ω given)	$\sum_{t=1}^{T} [y_t - f(x_t,\theta)]' \, \Omega^{-1} \, [y_t - f(x_t,\theta)]$

For the univariate and multivariate normal families, the P.M.L.E. are respectively the nonlinear least squares estimator and a minimum distance estimator (MALINVAUD, E. (1979)).

THEOREM 2 : A necessary condition for the P.M.L.E. associated with a family $\ell(u,m)$ $m \in M$ (the closure of an open connex set) to be strongly consistent for any θ , f , λ_0 satisfying $\mathcal{A}$ is that $\ell(u,m)$ is an exponential family of type I.

Proof : See Appendix 1

THEOREM 3 : If ($\ell(u,m)$, $m \in M$) is an exponential family of type I :

$$\sqrt{T}\,(\hat{\theta}_T - \theta_0) \xrightarrow{d} N(0, J^{-1}\, I\, J^{-1})$$

where :

$$J = E_x \left[\frac{\partial f}{\partial \theta} \frac{\partial C}{\partial m} \frac{\partial f'}{\partial \theta}\right] = E_x \left[\frac{\partial f}{\partial \theta} \Sigma^{-1} \frac{\partial f'}{\partial \theta}\right]$$

$$I = E_x \left(\frac{\partial f}{\partial \theta} \frac{\partial C}{\partial m} \Sigma_0 \frac{\partial C'}{\partial m} \frac{\partial f'}{\partial \theta}\right)$$

$$= E_x \left(\frac{\partial f}{\partial \theta} \Sigma^{-1} \Sigma_0 \Sigma^{-1} \frac{\partial f'}{\partial \theta}\right)$$

Σ_0 being the covariance matrix of λ_0 and $\frac{\partial f}{\partial \theta}$, Σ being evaluated at $\theta = \theta_0$.

Proof : The asymptotic normality is a consequence of the results mentioned in the preliminaries. The expressions of J and I are derived in Appendix 3.

5. QUASI-GENERALIZED PSEUDO-MAXIMUM LIKELIHOOD (Q.G.P.M.L.)

Property 5 : The set of asymptotic covariance matrices of $\hat{\theta}_T$ has a lower bound :

$$\forall\ \ell(u,m) \quad : \quad J^{-1} I J^{-1} \gg \mathcal{J} = \left[E_x \left(\frac{\partial f}{\partial \theta} \Sigma_0^{-1} \frac{\partial f'}{\partial \theta}\right)\right]^{-1}$$

where $\gg$ is the usual order relation on symmetric matrices.

Proof : $J^{-1}IJ^{-1} - \left[\underset{X}{E}\left(\frac{\partial f}{\partial\theta}\ \Sigma_o^{-1}\ \frac{\partial f'}{\partial\theta}\right)\right]^{-1}$

$$= \underset{X}{E}\left\{\left[J\ \frac{\partial f}{\partial\theta}\ \Sigma_o^{-1}\ \Sigma^{+\frac{1}{2}} - J^{-1}\ \frac{\partial f}{\partial\theta}\ \Sigma^{-\frac{1}{2}}\right]\ \Sigma^{-\frac{1}{2}}\ \Sigma_o\ \Sigma^{-\frac{1}{2}}\right.$$

$$\left.\left[\Sigma^{+\frac{1}{2}}\ \Sigma_o^{-1}\ \frac{\partial f'}{\partial\theta}\ J - \Sigma^{-\frac{1}{2}}\ \frac{\partial f'}{\partial\theta}\ J^{-1}\right]\right\} \gg 0$$

Q.E.D.

As it is shown hereafter the lower bound can be achieved by a two step estimation procedure.

As shown in table 2, some classical exponential families of type I (negative binomial, gamma, normal) also depend on an additional parameter η . This parameter is a function of m and Σ : $\eta = \Psi(m,\Sigma)$, where Ψ defines, for any given m , a one to one relationship between η and Σ . Therefore it is interesting to consider the more general class of distributions :

$$\ell^*(u,m,\eta) = \exp\{A(m,\eta) + B(\eta,u) + C(m,\eta)u\}$$

THEOREM 4 : Let $\tilde{\eta}_T(x)$ be a strongly consistent estimator of $\Psi[f(x,\theta_o),\Sigma_o(x)]$, the QGPML estimator of θ , obtained by maximizing :
$\underset{\theta}{\text{Max}} \sum_{t=1}^{T} \text{Log}\ \ell^*[y_t,f(x_t,\theta),\tilde{\eta}_T(x)]$, is under α strongly consistent and is asymptotically normal :
$\sqrt{T}\ (\hat{\theta}_T - \theta_o) \xrightarrow{d} N(o;J)$

Proof : Theorem 4 is a consequence of Burguete-Gallant's results, when a nuisance parameter is replaced by a strongly consistent estimator $\hat{\eta}_T$. The asymptotic covariance matrix of this estimator is the same as

the asymptotic covariance matrix of the PML estimator associated with $\ell(u,m) = \ell^*(u,m,\Psi[f(x,\theta_0),\sum_0(x)])$. In this case the matrix $\dot{\sum}$, evaluated at $m_0 = f(x,\theta_0)$, is equal to :
$\frac{\partial C}{\partial m}\{f(x,\theta_0),\Psi[f(x,\theta_0),\sum_0]\} = \sum_0$ and the asymptotic matrix is :

$$J^{-1}IJ^{-1} = \left[E_X\left(\frac{\partial f}{\partial\theta}\sum_0^{-1}\frac{\partial f'}{\partial\theta}\right)\right]^{-1} E_X\left(\frac{\partial f}{\partial\theta}\sum_0^{-1}\sum_0\sum_0^{-1}\frac{\partial f'}{\partial\theta}\right)\left[E_X\left(\frac{\partial f}{\partial\theta}\sum_0^{-1}\frac{\partial f'}{\partial\theta}\right)\right]^{-1} = J .$$

Q.E.D.

As an example let us consider the following model

$$y_t = f(x_t,\theta_0) + e_t \quad \text{with} \quad V(y_t|x_t) = g(x_t,\theta_0)^2\, \eta_0$$

In a first step, θ_0 may be consistently estimated by the nonlinear least squares estimator $\tilde{\theta}_T$ and a consistent estimate of η_0 is

$$\tilde{\eta}_T = \frac{1}{T}\sum_{t=1}^{T}\frac{[y_t - f(x_t,\tilde{\theta}_T)]^2}{g^2(x_t,\tilde{\theta}_T)} .$$

In a second step, it is possible to adopt the QGPML approach using for instance either the normal family or the gamma family. In the first case, the QGPML of θ is obtained by minimizing

$$\sum_{t=1}^{T}\frac{[y_t - f(x_t,\theta)]^2}{g(x_t,\tilde{\theta}_T)^2\,\tilde{\eta}_T},$$ which is simply the quasi generalized nonlinear least squares method. In the second case, the QGPML of θ is obtained by maximizing :

$$\sum_{t=1}^{T} - \tilde{\alpha}_T(x_t)\left[\log f(x_t,\theta) + \frac{y_t}{f(x_t,\theta)}\right] = \sum_{t=1}^{T} - \frac{f^2(x_t,\tilde{\theta}_T)}{g^2(x_t,\tilde{\theta}_T)\tilde{\eta}_T}\left[\log f(x_t,\theta) + \frac{y_t}{f(x_t,\theta)}\right]$$

These two estimators are asymptotically equivalent and reached the lower bound.

In the particular case $g(x_t,\theta) = 1$, the QGPML estimator based on the normal family is equal to the PML estimator. Similarly if $g(x_t,\theta) = f(x_t,\theta)$ the QGPML estimator and the PML estimator based on the gamma family are identical.

An other illustration of theorem 4 can be found in dichotomous qualitative response model, with repeated observations. The observed dependent variables y_{hi} $(h=1,\dots,H \; ; \; i=1,\dots,n_h)$ are independent and are such that : $P(y_{hi}=1) = F(x_h \theta)$, $P(y_{hi}=0) = 1 - F(x_h \theta)$. The properties of the estimators are discussed, when $n_h = \lambda_h n$, $n \to \infty$. The ML estimator of θ , which is a PML estimator based on the BERNOULLI distribution has an asymptotic covariance matrix equal to $\mathcal{J}$. In effect, since it is a PML estimator, its covariance matrix is greater than $\mathcal{J}$, and it is equal to $\mathcal{J}$, since this estimator is asymptotically efficient. The minimum chi-square estimator obtained by minimizing :

$$\sum_{h=1}^{H} \sum_{i=1}^{n_h} \frac{[y_{hi} - F(x_h \theta)]^2}{\bar{y}_h (1-\bar{y}_n)} \text{ , where } \bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}$$ is a QGPLM estimator

based on the normal distribution. Therefore this estimator is asymptotically efficient.

A last property of the QGPML estimators is the asymptotic efficiency, when the "true density function" belongs to the family :

$\lambda^*(y_t,x_t,\theta,\alpha) = \exp\{A^*[m(x_t),\eta(x_t)] + B^*[\eta(x_t),y_t] + C^*[m(x_t),\eta(x_t)]y_t\}$

with $m(x_t) = f(x_t,\theta)$, $\eta(x_t) = g(x_t,\alpha)$ and θ,α functionnally independent. This true density function λ_0^* is obtained for $\theta = \theta_0$ and $\alpha = \alpha_0$.

THEOREM 5 : The QGPML estimator of θ_0 is asymptotically equivalent to the maximum likelihood estimator of θ_0 obtained by maximizing (with respect to θ and α) : $\sum_{t=1}^{T} \log \lambda^*(y_t,x_t,\theta,\alpha)$.

Proof : Since the ML estimator $\begin{pmatrix}\theta_T^* \\ \alpha_T^*\end{pmatrix}$ of the parameters are consistent and asymptotically normal, we only have to compare $\mathcal{J}$ with the asymptotic covariance matrix of θ_T^* .

This matrix is given by :

$$V_{as}\ \sqrt{T}\begin{pmatrix}\theta_T^* - \theta_0 \\ \alpha_T^* - \alpha_0\end{pmatrix} = \left[\underset{x}{E}\ \underset{0}{E}\left(-\frac{\partial^2 \log \lambda^*(y,x,\theta_0,\alpha_0)}{\partial(\theta,\alpha)\ \partial(\theta,\alpha)'}\right)\right]^{-1}$$

Let us now show that θ_T^* and α_T^* are asymptotically independent ; we have :

$$\underset{x}{E}\ \underset{0}{E}\left(-\frac{\partial^2 \log \lambda^*(y,x,\theta_0,\alpha_0)}{\partial\theta\ \partial\alpha'}\right)$$

$$= \underset{x}{E}\left(\frac{\partial f}{\partial\theta}\ E_0\left(\frac{\partial A^*}{\partial m \partial \eta'} + y\ \frac{\partial C^*}{\partial m \partial \eta'}\right)\ \frac{\partial g'}{\partial\alpha}\right)$$

$$= \underset{x}{E}\left[\frac{\partial f}{\partial\theta}\left(\frac{\partial A^*}{\partial m \partial \eta'} + m_0\ \frac{\partial C^*}{\partial m \partial \eta'}\right)\ \frac{\partial g'}{\partial\alpha}\right]$$

where all the expressions are evaluated at the true values θ_o , α_o .
It is deduced from property 1 that $\frac{\partial A^*}{\partial m \partial \eta'} + m_o \frac{\partial C^*}{\partial m \partial \eta'} = 0$ and the asymptotic independence follows.

Therefore : $V_{as} [\sqrt{T}(\theta_T^* - \theta_o)] = \left\{ \underset{X}{E} \underset{o}{E} \left[- \frac{\partial^2 \log \lambda^*(y,x,\theta_o,\alpha_o)}{\partial\theta \, \partial\theta'} \right] \right\}^{-1}$

$$= \left\{ \underset{X}{E} \underset{o}{E} \left(\frac{\partial \text{ Log } \lambda^*}{\partial\theta} \frac{\partial \text{ Log } \lambda^{*\prime}}{\partial\theta} \right) \right\}^{-1}$$

$$= \left\{ \underset{X}{E} \left[\frac{\partial f}{\partial\theta} \underset{o}{E} \left(\frac{\partial C^*}{\partial m} (y-m)(y-m)' \frac{\partial C^{*\prime}}{\partial m} \right) \frac{\partial f'}{\partial\theta} \right] \right\}^{-1}$$

$$= \left\{ \underset{X}{E} \left[\frac{\partial f}{\partial\theta} \Sigma_o^{-1} \underset{o}{E} (y-m)(y-m') \Sigma_o^{-1} \frac{\partial f'}{\partial\theta} \right] \right\}^{-1}$$

$$= \left\{ \underset{X}{E} \left[\frac{\partial f}{\partial\theta} \Sigma_o^{-1} \frac{\partial f}{\partial\theta'} \right] \right\}^{-1} = \mathcal{J}$$

Q.E.D.

6. EXPONENTIAL FAMILIES OF TYPE II

The approach of sections 3 and 4 may be generalized to take into account the second order moments.

The model, which is now considered, is the following :

$y_t = f(x_t,\theta_o) + e_t$, where $V(e_t|x_t) = \Sigma_o(x_t) = g(x_t,\theta_o)$.

The parameter θ is assumed to be second order identifiable :

$$\left\{ \begin{array}{l} f(x,\theta_1) = f(x,\theta_o) \\ \\ g(x,\theta_1) = g(x,\theta_o) \end{array} \right. \quad \mu \text{ a.e.} \Longrightarrow \theta_o = \theta_1 .$$

The other assumptions of section 2 are satisfied.

Définition 2 : A family of probability measures on $\mathbb{R}^G$ indexed by $m \in M \subset \mathbb{R}^G$ and $\Sigma \in E$, where E is a subset of the positive definite matrices, is called exponential of type II, if :

a) every element of the family has a density function with respect to a given measure $\nu(du)$, which may be written as :

$\ell(u,m,\Sigma) = \exp\{A(m,\Sigma) + B(u) + C(m,\Sigma)u + u' D(m,\Sigma)u\}$

$A(m,\Sigma)$, $B(u)$ are scalar, $C(m,\Sigma)$ is a row vector of size G and $D(m,\Sigma)$ is a square matrix (G,G).

b) m is the mean and Σ is the covariance matrix of the distribution $\ell(u,m,\Sigma)$. These parameters are identifiable.

We also assume that conditions $\mathcal{A}$ are satisfied for f, g such that $f(x,\theta) \in M$ $\forall x$, $\forall \theta$, $g(x,\theta) \in E$ $\forall x$, $\forall \theta$ and for φ defined by :

$$\varphi(y,x,\theta) = \text{Log } \ell[y,f(x,\theta),g(x,\theta)]$$

Examples of such families are for instance the normal distribution $N(m,\Sigma)$ or the discrete distribution on $\{-1,0,1\}$, with probabilities p_1,p_2,p_3 ; in the latter case $\ell(u,m,\sigma^2) = p_1^{\frac{u(1+u)}{2}} \; p_2^{1-u^2} \; p_3^{\frac{u(-1+u)}{2}}$ with $p_1 + p_2 + p_3 = 1$, $p_1 - p_3 = m$, $p_1 + p_3 = \sigma^2 + m^2$.

Property 6 : $\forall\, m,m_o \in M$ $\forall\, \Sigma,\Sigma_o \in E$:

$$A(m,\Sigma) + C(m,\Sigma)m_o + \text{Tr } D(m,\Sigma)(\Sigma_o + m_o m_o')$$
$$\leq A(m_o,\Sigma_o) + C(m_o,\Sigma_o)m_o + \text{Tr } D(m_o,\Sigma_o)(\Sigma_o + m_o m_o')$$

The equality is possible, if and only if $m = m_o$, $\Sigma = \Sigma_o$.

Proof : This result is, as for property 4, a direct consequence of KULLBACK's inequality.

Q.E.D.

From this property and the second order identifiability condition, we deduce, as for theorem 1 and 2, the strong consistency and the asymptotic normality of the PML estimator of θ based on $\ell(u,m,\Sigma)$.

THEOREM 6 : If $(\ell(u,m,\Sigma)\ \ m \in M\ ,\ \Sigma \in E\ )$ is an exponential family of type II, the estimator of θ obtained by maximizing :

$$\underset{\theta}{\text{Max}} \sum_{t=1}^{T} \text{Log}\ell\ [y_t, f(x_t,\theta), g(x_t,\theta)]$$ is strongly consistent and asymptotically normal.

The proof is similar to that of theorem 1.

However the expression of the asymptotic covariance matrix is rather complicated (see appendix 4).

It is also possible to obtain the reciprocal of the previous theorem, in the case $G = 1$.

THEOREME 7 : A necessary condition for the PML estimator associated with a family $\ell(u,m,\sigma^2)$ $m \in M$, $\sigma^2 \in E$ (M and E are closures of open connex sets) to be strongly consistent for any θ , f , g , λ_0 satisfying $\mathcal{A}$ is that $\ell(u,m,\sigma^2)$ is an exponential family of type II.

Proof : See Appendix 2

7. CONCLUSION

The results obtained in this paper are clearly related to the general problem of robustness and could be used in many contexts. This is the reason of the high level of generality kept throughout the paper. A companion paper (GOURIEROUX, MONFORT, TROGNON (1981)) describes the application of these estimation methods to econometric models for discrete data.

---==ooOoo==---

APPENDIX 1

Proof of theorem 2 :

We are going to show that the condition is already necessary, if we consider the special case, in which the y_t are i.i.d , with mean $\theta_o \in \Theta = M \subset \mathbb{R}^G$.

First step :

The PMLE exists since Θ is compact and φ_T is continuous on Θ . If this estimator $\hat{\theta}_T$ is strongly consistent of θ_o , we deduce from the strong uniform convergence of $\varphi_T(y^T,x^T,\theta)$ to $\varphi_\infty(\theta)$, that θ_o provides the maximum of φ_∞ . Since $\theta_o \in \overset{o}{\Theta}$ and since φ_∞ is differentiable, we have :

$$\frac{d\varphi_\infty}{d\theta}(\theta_o) = \underset{X}{E}\,\underset{o}{E} \frac{\partial \text{ Log } \ell(y,\theta_o)}{\partial\theta} = \underset{o}{E} \frac{\partial \text{ Log } \ell(y,\theta_o)}{\partial\theta} = 0$$

Second step :

Let us first consider the case $G = 1$.

For any distribution λ_o such that $\underset{o}{E}\, y = \theta_o$ we have $\underset{o}{E} \dfrac{\partial \text{ Log } \ell(y,\theta_o)}{\partial\theta} = 0$.

In particular, this is true if the support of λ_o consists in two points y_1 and y_2 , with $y_1 < \theta_o < y_2$; therefore :

$$\forall\, p_1 , p_2 \begin{cases} p_1 + p_2 = 1 \\ \\ p_1 y_1 + p_2 y_2 = \theta_o \end{cases} \Longrightarrow p_1 \frac{\partial \text{ Log } \ell(y_1,\theta_o)}{\partial\theta} + p_2 \frac{\partial \text{ Log } \ell(y_2,\theta_o)}{\partial\theta} = 0$$

This implies that for any y_1, y_2 such that $y_1 < \theta_o < y_2$

$$(y_2 - \theta_o) \frac{\partial \text{ Log } \ell(y_1,\theta_o)}{\partial\theta} + (\theta_o - y_1) \frac{\partial \text{ Log } \ell(y_2,\theta_o)}{\partial\theta} = 0$$

$$\frac{\dfrac{\partial \text{ Log } \ell(y_1,\theta_o)}{\partial\theta}}{y_1 - \theta_o} = \frac{\dfrac{\partial \text{ Log } \ell(y_2,\theta_o)}{\partial\theta}}{y_2 - \theta}$$

Considering y_2 as a variable and y_1 as fixed and then allowing y_1 to vary, it is seen that there exists a scalar $\lambda(\theta_o)$ such that $\frac{\partial \text{ Log } \ell(y,\theta_o)}{\partial\theta} = \lambda(\theta_o) \cdot (y - \theta_o)$. The result follows by integrating both sides with respect to θ in $\mathring{M} = \mathring{\Theta}$ and extending to Θ by continuity arguments (see $\mathcal{a}$).

In the general G-dimensional case; the same kind of proof can be developed by considering distributions whose supports consists in G+1 independent points admitting θ_o as a barycenter.

APPENDIX II

Proof of theorem 7 (case $G=1$)

We shall see that the condition of theorem 7 is already necessary if we restrict ourselves to the case in which the y_t are i.i.d. with mean $f(x,\theta_o) = m_o$ and variance $g(x_t,\theta_o) = \sigma_o^2$.

The same arguments as those given in the first step of appendix I imply that

$$\underset{o}{E} \frac{\partial \text{ Log } \ell(y,m,\sigma^2)}{\partial m} = 0$$

and

$$\underset{o}{E} \frac{\partial \text{ Log } \ell(y,m,\sigma^2)}{\partial \sigma^2} = 0$$

In particular, this is true if the support of λ_o consists in three points $y_1 < y_2 = m_o < y_3$; therefore :

$$\forall\ p_1,p_2,p_3 : (1) \left\{ \begin{array}{l} p_1 + p_2 + p_3 = 1 \\ p_1y_1 + p_2m_o + p_3y_3 = m_o \\ p_1y_1^2 + p_2m_o^2 + p_3y_3^2 = m_o^2 + \sigma_o^2 \end{array} \right.$$

$$(2) \left\{ \begin{array}{l} p_1 \dfrac{\partial \text{Log}\ell(y_1,m_o,\sigma_o^2)}{\partial m} + p_2 \dfrac{\partial \text{Log}(m_o,m_o,\sigma_o^2)}{\partial m} + p_3 \dfrac{\partial \text{Log}(y_3,m_o,\sigma_o^2)}{\partial m} = 0 \\ p_1 \dfrac{\partial \text{Log}\ell(y_1,m_o,\sigma_o^2)}{\partial \sigma^2} + p_2 \dfrac{\partial \text{Log}(m_o,m_o,\sigma_o^2)}{\partial \sigma^2} + p_3 \dfrac{\partial \text{Log}(y_3,m_o,\sigma_o^2)}{\partial \sigma^2} = 0 \end{array} \right.$$

Denoting $\dfrac{y_j - m_o}{\sigma_o^2}$ by y_j^* , system (1) becomes

$$\left\{ \begin{array}{l} p_1 + p_2 + p_3 = 1 \\ p_1y_1^* + p_3y_3^* = 0 \\ p_1y_1^{*2} + p_3y_3^{*2} = 1 \end{array} \right. \qquad \text{with } y_1^* < 0 < y_3^*$$

Solving this system, we obtain

$$p_1 = \frac{-1}{\overset{*}{y}_1(\overset{*}{y}_3 - \overset{*}{y}_1)} \quad , \quad p_2 = \frac{\overset{*}{y}_1\overset{*}{y}_3 + 1}{\overset{*}{y}_1\,\overset{*}{y}_3} \quad , \quad p_3 = \frac{-1}{\overset{*}{y}_3(\overset{*}{y}_1 - \overset{*}{y}_3)}$$

This solution is a probability distribution if, and only if, $\overset{*}{y}_1\overset{*}{y}_3 \leq -1$.

For such a probability distribution we have :

$$\left\{\begin{array}{l} p_1\overset{*}{y}_1 + p_3\overset{*}{y}_3 = 0 \\ p_1(\overset{*2}{y_1} - 1) - p_2 + p_3(\overset{*2}{y_3} - 1) = 0 \\ p_1 \dfrac{\partial \mathrm{Log}\ell^*(\overset{*}{y}_1,m_0,\sigma_0^2)}{\partial m} + p_2 \dfrac{\partial \mathrm{Log}\ell^*(\overset{*}{y}_2,m_0,\sigma_0^2)}{\partial m} + p_3 \dfrac{\partial \mathrm{Log}\ell^*(\overset{*}{y}_3,m_0,\sigma_0^2)}{\partial m} = 0 \end{array}\right.$$

with $\dfrac{\partial \mathrm{Log}\ell^*}{\partial m}(\overset{*}{y}_i,m_0,\sigma_0^2) = \dfrac{\partial \mathrm{Log}\ell}{\partial m}(y_i,m_0,\sigma_0^2)$

Since this homogeneous system has a non-zero solution p_1,p_2,p_3 , this implies

$$\Delta(\overset{*}{y}_1,\overset{*}{y}_3) = \det \begin{vmatrix} \overset{*}{y}_1 & 0 & \overset{*}{y}_3 \\ \overset{*2}{y_1} - 1 & -1 & \overset{*2}{y_3} - 1 \\ \dfrac{\partial \mathrm{Log}\ell^*(\overset{*}{y}_1,m_0,\sigma_0^2)}{\partial m} & \dfrac{\partial \mathrm{Log}\ell^*(0,m_0,\sigma_0^2)}{\partial m} & \dfrac{\partial \mathrm{Log}\ell^*(\overset{*}{y}_3,m_0,\sigma_0^2)}{\partial m} \end{vmatrix}$$

$\forall \ \overset{*}{y}_1,\overset{*}{y}_3 \in \mathcal{D} = \{\overset{*}{y}_1 < 0 < \overset{*}{y}_3 \ , \ \overset{*}{y}_1\overset{*}{y}_3 \leq -1\}$

This also implies that, for any $(\overset{*}{y}_1,\overset{*}{y}_3)$ belonging to $\mathcal{D}$, the functions $\dfrac{\partial \mathrm{Log}\ell^*}{\partial m}(\overset{*}{y}_1,m_0,\sigma_0^2)$ and $\dfrac{\partial \mathrm{Log}\ell^*}{\partial m}(\overset{*}{y}_3,m_0,\sigma_0^2)$ can be written as :

$$\left\{\begin{array}{l} \dfrac{\partial \mathrm{Log}\ell^*(\overset{*}{y}_3,m_0,\sigma_0^2)}{\partial m} = \alpha(\overset{*}{y}_1) + \beta(\overset{*}{y}_1)\overset{*}{y}_3 - \alpha(\overset{*}{y}_1)\overset{*2}{y_3} \\ \dfrac{\partial \mathrm{Log}\ell^*(\overset{*}{y}_1,m_0,\sigma_0^2)}{\partial m} = \tilde{\alpha}(\overset{*}{y}_3) + \tilde{\beta}(\overset{*}{y}_3)\overset{*}{y}_1 - \tilde{\alpha}(\overset{*}{y}_3)\overset{*2}{y_1} \end{array}\right.$$

Considering successively y_1^* as a variable and y_3^* fixed and then y_3^* as a variable and y_1^* fixed, it is seen that the functions α,β , $\tilde{\alpha},\tilde{\beta}$ are constant ; therefore we have :

$$\begin{cases} \dfrac{\partial \mathrm{Log}\ell^*}{\partial m}(y_3^*,m_0,\sigma_0^2) = \alpha + \beta y_3^* - \alpha y_3^{*2} & \forall\ y_3^* > 0 \\ \dfrac{\partial \mathrm{Log}\ell^*(y_1^*,m_0,\sigma_0^2)}{\partial m} = \tilde{\alpha} + \tilde{\beta} y_1^* - \tilde{\alpha} y_1^{*2} & \forall\ y_1^* < 0 \end{cases}$$

Replacing in the equation $\Delta(y_1^*,y_3^*) = 0$, $\frac{\partial \mathrm{Log}\ell^*}{\partial m}(y_3^*,m_0,\sigma_0^2)$ and $\frac{\partial \mathrm{Log}\ell^*}{\partial m}(y_1^*,m_0,\sigma_0^2)$ by the previous expressions we obtain :

$$\frac{\partial \mathrm{Log}\ell^*(0,m_0,\sigma_0^2)}{\partial m} = \alpha = \tilde{\alpha}$$

and $\beta = \tilde{\beta}$.

It follows that

$$\frac{\partial \mathrm{Log}\ell^*(y^*,m,\sigma^2)}{\partial m} = \alpha(m,\sigma^2) + \beta(m,\sigma^2)y^* - \alpha(m,\sigma^2)y^{*2} \qquad \forall\ m,\sigma^2,y^*$$

$$\frac{\partial \mathrm{Log}\ell(y,m,\sigma^2)}{\partial m} = a(m,\sigma^2) + c(m,\sigma^2)y + d(m,\sigma^2)y^2 \qquad \forall\ m,\sigma^2,y$$

Integrating with respect to m , we get :

(3) $\qquad \mathrm{Log}\ell(y,m,\sigma^2) = A(m,\sigma^2) + B(\sigma^2,y) + C(m,\sigma^2)y + D(m,\sigma^2)y^2$

The same argument applied to σ^2 instead of m gives :

(4) $\qquad \mathrm{Log}\ell(y,m,\sigma^2) = A^*(m,\sigma^2) + B^*(m,y) + C^*(m,\sigma^2)y + D^*(m,\sigma^2)y^2$

The previous equation shows that the terms depending simultaneously on y and σ^2 are quadratic in y , and from (3) we obtain

$$\mathrm{Log}\ell(y,m,\sigma^2) = A(m,\sigma^2) + B(y) + C(m,\sigma^2)y + D(m,\sigma^2)y^2$$

APPENDIX 3

COMPUTATION OF THE ASYMPTOTIC COVARIANCE MATRIX OF THE PSEUDO-MAXIMUM LIKELIHOOD ESTIMATOR IN THE CASE OF EXPONENTIAL FAMILY OF TYPE I .

As it is proved by BURGUETE-GALLANT, the asymptotic variance-covariance matrix of the PMLE is given by :

$$V = J^{-1} I J^{-1}$$

where

$$J = \underset{X}{E} \underset{o}{E} \left[- \frac{\partial^2 \text{Log}\ell}{\partial\theta \, \partial\theta'} (y, f(x,\theta_o)) \right]$$

$$I = \underset{X}{E} \underset{o}{E} \left[\frac{\partial \text{Log}\ell}{\partial\theta} (y, f(x,\theta_o)) \frac{\partial \text{Log}\ell}{\partial\theta} (y, f(x,\theta_o))' \right]$$

If ℓ is an element of an exponential family of type I, we have :

(A2-1) $$\frac{\partial \text{Log}\ell}{\partial\theta}(y,x,\theta) = \frac{\partial}{\partial\theta} [A(f(x,\theta)) + B(y) + C(f(x,\theta))y]$$

$$= \frac{\partial f(x,\theta)}{\partial\theta} \left(\frac{\partial A}{\partial m} (f(x,\theta)) + \frac{\partial C}{\partial m} (f(x,\theta))y \right)$$

$$= \frac{\partial f}{\partial\theta} \left(\frac{\partial A}{\partial m} + \frac{\partial C}{\partial m} y \right)$$

$$= \frac{\partial f}{\partial\theta} \frac{\partial C}{\partial m} (y-m) \qquad \text{(property 1)}$$

and

$$\text{(A2.2)}\quad \frac{\partial^2 \text{Log}\ell}{\partial\theta\ \partial\theta^{T}}(y,f(x,\theta)) = \frac{\partial}{\partial\theta}\left[\sum_g \frac{\partial f_g}{\partial\theta}\left(\frac{\partial A}{\partial m_g} + \frac{\partial C}{\partial m_g}y\right)\right]'$$

$$= \sum_g \frac{\partial^2 f_g}{\partial\theta\partial\theta^{T}}\left(\frac{\partial A}{\partial m_g} + \frac{\partial C'}{\partial m_g}y\right) + \frac{\partial f}{\partial\theta}\left[\frac{\partial^2 A}{\partial m\partial m^{T}} + \sum_g \frac{\partial^2 C_g}{\partial m\ \partial m^{T}} y_g\right]\frac{\partial f'}{\partial\theta}$$

$$= \sum_g \frac{\partial^2 f_g}{\partial\theta\ \partial\theta^{T}}\frac{\partial C'}{\partial m_g}(y-m) + \frac{\partial f}{\partial\theta}\left[\sum_g \frac{\partial^2 C_g}{\partial m\ \partial m^{T}}(y_g-m_g) - \frac{\partial C}{\partial m}\right]\frac{\partial f'}{\partial\theta}$$

(properties 1 and 2)

From (A2-1) we deduce :

$$I = \underset{x}{E}\left[\frac{\partial f}{\partial\theta}\frac{\partial C}{\partial m}\ \underset{o}{E}\left[(y-m)(y-m)'\right]\frac{\partial C'}{\partial m}\ \frac{\partial f'}{\partial\theta}\right] = \underset{x}{E}\left[\frac{\partial f}{\partial\theta}\frac{\partial C}{\partial m}\sum_o \frac{\partial C'}{\partial m}\frac{\partial f'}{\partial\theta}\right]$$

From (A2.2) it follows

$$J = -\underset{x}{E}\left\{\frac{\partial f}{\partial\theta}\left[\sum_g \frac{\partial^2 C_g}{\partial m\ \partial m^{T}}\ \underset{o}{E}\ (y_g - f_g(x,\theta_o)) - \frac{\partial C}{\partial m}\right]\frac{\partial f'}{\partial\theta}\right\}$$

$$= +\underset{x}{E}\left[\frac{\partial f}{\partial\theta}\frac{\partial C}{\partial m}\frac{\partial f'}{\partial\theta}\right]$$

where $\frac{\partial f}{\partial\theta}$, $\frac{\partial C}{\partial m}$ stand for : $\frac{\partial f}{\partial\theta}(x,\theta_o)$, $\frac{\partial C}{\partial m}(f(x,\theta_o))$ respectively.

APPENDIX 4

COMPUTATION OF THE ASYMPTOTIC COVARIANCE MATRIX OF THE PSEUDO-MAXIMUM LIKELIHOOD ESTIMATOR IN THE CASE OF EXPONENTIAL FAMILIES OF TYPE II . (G=1)

Consider the following exponential family of type II :

$\ell(u,m,\sigma^2) = \exp[A(m) + B(u) + C(m,\sigma^2)u + D(m,\sigma^2)u^2]$ $(u \in \mathbb{R})$

ℓ is a density function, then : $\int \ell(u,m,\sigma^2)\, \nu(du) = 1$. Deriving this last equality with respect to $\alpha' = (m,\sigma^2)$ leads to :

$$\text{(A4.1)} \qquad \frac{\partial A}{\partial \alpha} + \frac{\partial C}{\partial \alpha} m + \frac{\partial D}{\partial \alpha}(m^2+\sigma^2) = 0$$

Since the mean and the variance of ℓ are m and σ^2 , it follows :

$$\text{(A4.2)} \qquad \int u\, \ell(u,m,\sigma^2)\, \nu(du) = m$$

$$\text{(A4.3)} \qquad \int u^2\, \ell(u,m,\sigma^2)\, \nu(du) = m^2 + \sigma^2 .$$

The derivation of (A4.2) which respect to $\alpha' = (m,\sigma^2)$ implies :

$$\text{(A4.4)} \qquad \frac{\partial A}{\partial \alpha} m + \frac{\partial C}{\partial \alpha}(m^2+\sigma^2) + \frac{\partial D}{\partial \alpha} E\, u^3 = \binom{1}{0}$$

The derivation of (A4.3) with respect to α yields to :

$$\text{(A4.5)} \qquad \frac{\partial A}{\partial \alpha}(m^2+\sigma^2) + \frac{\partial C}{\partial \alpha} E\, u^3 + \frac{\partial D}{\partial \alpha} E\, u^4 = \begin{pmatrix} 2\, m \\ 1 \end{pmatrix}$$

the computation of the asymptotic covariance matrix of the PMLE is based on

$$J = \underset{x}{E}\,\underset{o}{E} - \frac{\partial^2 \text{Log}\ell}{\partial\theta\,\partial\theta'}(g, f(x,\theta_0), g(x,\theta_0)) \text{ and}$$

$$I = \underset{x}{E}\,\underset{o}{E} \frac{\partial \text{Log}\ell}{\partial\theta}(y, f(x,\theta_0), g(x,\theta_0)) \frac{\partial \text{Log}\ell}{\partial\theta}(y, f(x,\theta_0), g(x,\theta_0))'$$

the first and second derivatives of $\text{Log}\ell$ are in that case :

$$\frac{\partial \text{Log}\ell}{\partial\theta} = \frac{\partial h}{\partial\theta}\left[\frac{\partial A}{\partial\alpha} + \frac{\partial C}{\partial\alpha} y + \frac{\partial D}{\partial\alpha} y^2\right] \qquad \text{with} \quad h = (f,g)$$

$$\frac{\partial^2 \text{Log}}{\partial\theta\,\partial\theta'} = \frac{\partial h}{\partial\theta}\left[\frac{\partial^2 A}{\partial\alpha\partial\alpha'} + \frac{\partial^2 C}{\partial\alpha\partial\alpha'} y + \frac{\partial^2 D}{\partial\alpha\partial\alpha'} y^2\right] \frac{\partial h'}{\partial\theta}$$

Let us consider first the conputation of J :

$$J = -\underset{x}{E}\,\underset{o}{E} \frac{\partial^2 \text{Log}\ell}{\partial\theta\,\partial\theta'} = -\underset{x}{E}\left[\frac{\partial f}{\partial\theta}\left[\frac{\partial^2 A}{\partial\alpha\partial\alpha'} + \frac{\partial^2 C}{\partial\alpha\partial\alpha'} f(x,\theta_0) + \frac{\partial^2 D}{\partial\alpha\partial\alpha'}(f(x,\theta_0)^2 + g(x,\theta_0))\right]\right]$$

For expository purpose we will denote $f(x,\theta_0)$ by m_0 and $g(x,\theta_0)$ by σ_0^2 . With these notations (A4.1) implies :

$$-\left\{\frac{\partial^2 A}{\partial\alpha\partial\alpha'} + \frac{\partial^2 C}{\partial\alpha\partial\alpha'} m_0 + \frac{\partial^2 D}{\partial\alpha\partial\alpha'}(m_0^2 + \sigma_0^2)\right\} = \frac{\partial C}{\partial\alpha}(1:0) + \frac{\partial D}{\partial\alpha}(2m_0:1)$$

Finally J is defined by :

$$\boxed{J = \underset{x}{E}\left\{\frac{\partial h}{\partial\theta}\left[\frac{\partial C}{\partial\alpha}(1:0) + \frac{\partial D}{\partial\alpha}(2m_0:1)\right]\frac{\partial h'}{\partial\theta}\right\}}$$

where $\frac{\partial C}{\partial\alpha}$, $\frac{\partial D}{\partial\alpha}$ are computed at $m_0 = f(x,\theta_0)$ and $\sigma_0^2 = g(x,\theta_0)$ and $\frac{\partial h}{\partial\theta}$ at θ_0 .

Let us now consider the computation of I which is defined by :

$$I = \underset{x}{E}\,\underset{o}{E}\left\{\frac{\partial \text{Log}\ell}{\partial\theta}(y, f(x,\theta_0), g(x,\theta_0)) \frac{\partial \text{Log}\ell}{\partial\theta}(y, f(x,\theta_0), g(x,\theta_0))'\right\}$$

Let $m_3^0 = \underset{o}{E}\, y^3$ and $m_4^0 = \underset{o}{E}\, y^4$ be the third and fourth product-moments of λ_0 the true distribution of the observations and m_3 and m_4 the similar moments for the family considered, evaluated at $m = m_0$, $\sigma^2 = \sigma_0^2$.

$$I = \underset{x}{E} \left\{ \frac{\partial h}{\partial \theta} \left\{ \frac{\partial C}{\partial \alpha} \left\{ \frac{\partial A}{\partial \alpha} m_o + \frac{\partial C}{\partial \alpha} (m_o^2 + \sigma_o^2) + \frac{\partial D}{\partial \alpha} m_3^0 \right\} + \frac{\partial D}{\partial \alpha} \left\{ \frac{\partial A}{\partial \alpha} (m_o + \sigma_o^2) + \frac{\partial C}{\partial \alpha} m_3^0 + \frac{\partial D}{\partial \alpha} m_4^0 \right\}' \right) \frac{\partial h'}{\partial \theta} \right\}$$

From (A4.4) and (A4.5) it follows :

$$I = \underset{x}{E} \left\{ \frac{\partial h}{\partial \theta} \left(\left[\frac{\partial C}{\partial \alpha} (1:0) + \frac{\partial D'}{\partial \alpha} (m_3^0 - m_3) \right] + \frac{\partial D}{\partial \alpha} \left[(2m_o:1) + \frac{\partial C'}{\partial \alpha} (m_3^0 - m_3) + \frac{\partial D'}{\partial \alpha} (m_4^0 - m_4) \right] \right) \frac{\partial h'}{\partial \theta} \right\}$$

APPENDIX 5

APPLICATION OF THE METHOD OF SCORING TO EXPONENTIAL FAMILIES OF TYPE I.

Let us examine the PML estimation problem with the following pseudo-likelihood function of type I :

$$\sum_{t=1}^{T} \text{Log}\,\ell(y_t, f(x_t b)) = \sum_{t=1}^{T} \{A(f(x_t b)) + B(y_t) + C(f(x_t b))\, y_t\}$$

where $A(.)$, $B(.)$ and $C(.)$ are scalar functions.

The pseudo maximum likelihood estimate may be obtained by the method of scoring ; the $(h+1)^{th}$ iteration is :

$$\tilde{b}_{h+1} = \tilde{b}_h - \left[E_o \sum_{t=1}^{T} \frac{\partial^2 \text{Log}\,\ell}{\partial b\, \partial b'} \right]^{-1}_{b=\tilde{b}_h} \left[\sum_{t=1}^{T} \frac{\partial \text{Log}\,\ell}{\partial b} \right]_{b=\tilde{b}_h}$$

$$= \tilde{b}_h + \left[\sum_t x_t x_t' \left(\frac{\partial f}{\partial m}\right)^2 \frac{\partial C}{\partial m} \right]^{-1} \left[\sum_t x_t' \frac{\partial f}{\partial m} \frac{\partial C}{\partial m} (y_t - f(x_t \tilde{b}_h)) \right]$$

where $\frac{\partial f}{\partial m}$ and $\frac{\partial C}{\partial m}$ are the derivatives of $f(m)$ and $c(m)$ evaluated at $m = x_t \tilde{b}_h$

No further computation program is needed to apply this method of scoring since it can easily be achieved by iterative GLS. In effect let us consider the nonlinear heteroscedastic pseudo-model defined by :

$$y_t = f(x_t b) + u_t\ , \quad E u_t = 0\ , \quad V u_t = \left(\frac{\partial C}{\partial m} (f(x_t b))\right)^{-1} \quad t=1,\dots,T$$

($V u_t$ is the variance associated with the pseudo distribution function)

$f(x_t b)$ may be expanded in a neighbourhood of $\tilde{b}_h$:

$$y_t \neq f(x_t \tilde{b}_h) + \frac{\partial f}{\partial m}(x_t \tilde{b}_h) \,.\, x_t (b - \tilde{b}_h) + u_t$$

This expression can also be written as :

$$y_t - f(x_t\tilde{b}_h) + \frac{\partial f}{\partial m}(x_t\tilde{b}_h)\, x_t\tilde{b}_h \neq \frac{\partial f}{\partial m}(x_t\tilde{b}_h)\, x_t b + u_t \qquad t = 1,\ldots,T$$

If we denote by $\hat{b}_{h+1}$ the GLS estimate of b obtained from this linear model with covariance matrix $\text{diag}\left\{\left(\frac{\partial C}{\partial m}(x_t\tilde{b}_h)\right)^{-1}\right\}$, it is straightforward to show that : $\tilde{b}_{h+1} = \hat{b}_{h+1}$

REFERENCES

BURGUETE, J. and GALLANT R. (1980) : "On unification of the asymptotic theory of nonlinear econometric models", C.P. 1296, North Carolina State University.

GALLANT, R. and HOLLY, A. (1980) : "Statistical inference in an implicit, nonlinear, simultaneous equation model in the context of maximum likelihood estimation". Econometrica 48, 697-720.

GOURIEROUX, C., MONFORT, A. and TROGNON, A. (1981): "Pseudo maximum likelihood methods : applications to Poisson models", INSEE D.P..

JENNRICH, R. (1969) : "Asymptotic properties of nonlinear least squares estimators", The Annals of Mathematical Statistics 40, 633-643.

JOHNSON, N. and KOTZ, S. (1969) : "Discrete distributions", Boston.

MALINVAUD, E. (1970) : "The consistency of nonlinear regressions", The Annals of Mathematical Statistics 41, 956-969.

MALINVAUD, E. (1979) : "Statistical methods of econometrics", Dunod, 3rd edition, 1979.

MONFORT, A. (1981) : "Cours de statistique mathématique", Economica, Paris.