OCTOBRE 1984.

N° 8502

SIMULATED RESIDUALS

GOURIEROUX, C.[*], MONFORT, A.[**], RENAULT, E.[***] TROGNON, A.[***]

* Université PARIS IX et CEPREMAP.

** INSEE.

*** ENSAE.

RESIDUS SIMULES

GOURIEROUX C., MONFORT A., RENAULT E.
TROGNON A.

Dans cet article on considère les modèles déduits d'un modèle de régression latent par une application non-linéaire (par exemple les modèles probit, tobit, de déséquilibre...). On définit, dans ce contexte, des résidus simulés qui jouent un rôle analogue à celui des résidus habituels dans les modèles de régression. Ces résidus simulés permettent en particulier d'utiliser les programmes de régression standard pour effectuer des vérifications graphiques ou des véritables tests statistiques.

SIMULATED RESIDUALS

GOURIEROUX C., MONFORT A., RENAULT E.
TROGNON A.

In this article we consider models deduced from a latent regression model by a non-linear mapping (probit, tobit, disequilibrium models...). In this context we define simulated residuals whose role is similar to that of usual residuals in the regression model. In particular it is possible, with these new tools, to use the standard regression packages for diagnostic checks or genuine statistical tests.

1. INTRODUCTION.

The large use of linear or nonlinear regression models in econometrics has two main origins : first the simplicity of the least squares method and, secondly, the existence of packages giving a number of diagnostic tools, which can be easily interpreted in a descriptive or in a statistical way. Such tools are, for instance, the residual plots, the sum of squares residuals, the R^2-coefficient, the Student or Fisher type statistics....

The main purpose of this paper is to explain how to use the standard regression packages for a large class of nonlinear models : models deduced from a latent regression model by a nonlinear mapping. This class contains as special cases the usual probit models, the simple or the generalised tobit models, the disequilibrium models....

The central idea of the present paper consists in simulating the values of the unobservable endogenous variables and in mechanically implementing the regression package on the simulated series.

In order to motivate this approach let us briefly consider an artificial probit model. The data generating process is :

$$y_i^* = 1 + x_i + u_i \ ,$$

where the u_i's are I.I.N. (0,1) and the x_i's are such that :

$$x_i = 4.899 \ 10^{-2} \ i - 3.474 \ .$$

The y_i^*'s are generated for $i = 1,\dots,100$; this implies that the empirical mean of the x_i's is -1, their empirical variance is 2 and the theoretical R^2 is .667 .

The observable endogenous variables are defined by :

$$\begin{cases} y_i = 1 & \text{if} \quad y_i^* > 0 \\ y_i = 0 & \text{if} \quad y_i^* < 0 \end{cases} .$$

In the probit model based on the linear regression :

$$y_i^* = \alpha + \beta x_i + u_i \qquad u_i \approx N(0,1) \ ,$$

the maximum likelihood estimators are :

$$\hat{\alpha} = \underset{(.193)}{.872} \ , \qquad \hat{\beta} = \underset{(.149)}{.891} \ .$$

It is now possible to simulate values z_i for the latent endogenous variables from the conditional distributions of y_i^* given y_i , in which α and β are replaced by $\hat{\alpha}$ and $\hat{\beta}$; this is easily done by taking z_i as the first positive (resp. negative) $z_{ij} = \hat{\alpha} + \hat{\beta} x_i + u_{ij}$, where the u_{ij} are $IIN(0,1)$, if y_i is equal to 1 (resp. equal to 0).

Using a standard package for regressing the z_i's on the x_i's , we obtain the following "second stage" results :

$$\tilde{\alpha} = \underset{(.119)}{.801} \ , \qquad \tilde{\beta} = \underset{(.068)}{.861} \ .$$

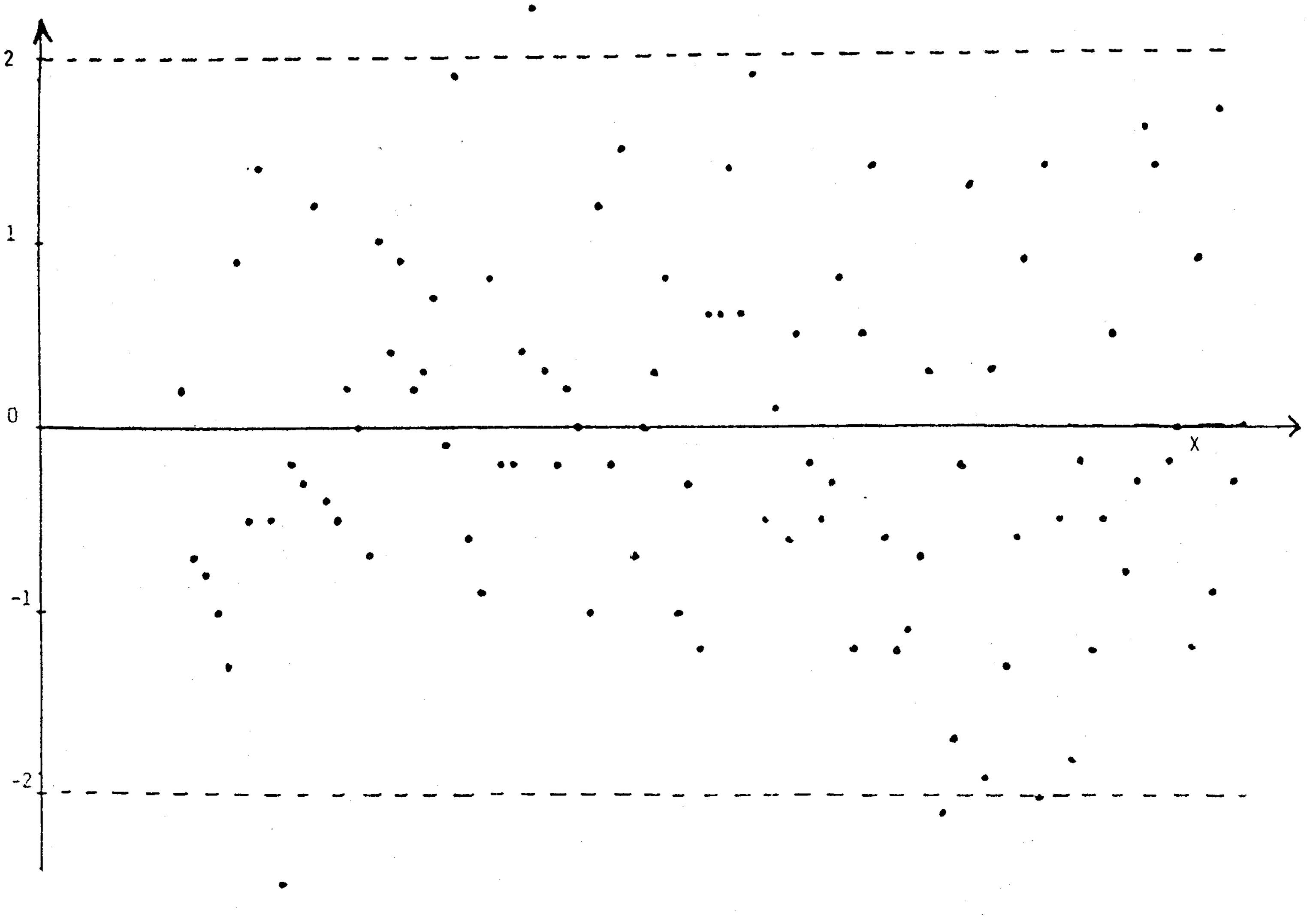
$$\text{M.S.E.} : .937 \qquad R^2 = .618 \ .$$

The residual plots have the following form :

FIGURE 1.

PLOT OF SIMULATED RESIDUALS

FOR A WELL SPECIFIED PROBIT MODEL



It is clear that this regression has the advantage to provide automatically a large number of statistical tools ; however it is important to keep in mind that the usual interpretations of these tools may be misleading, since they are based on simulated approximations of the y_i^*'s . For instance, we see, from the previous example, that $\tilde{\alpha}$ and $\tilde{\beta}$ are not too far from α and β , the R^2 and the M.S.E. are close to their theoretical values .667 and 1 .

The residual plots look similar to those obtained in the usual linear models. However the standard deviation of $\tilde{\alpha}$ and $\tilde{\beta}$ are obviously underestimated, since the computed values .119 and .068 are significantly smaller than the standard error of the asymptotically efficient estimators $\hat{\alpha}$, $\hat{\beta}$. This underestimation also shows that the tests based on the t or on the F statistics are not directly applicable.

In summary the previous approach is appealing from a descriptive point of view but, on the other hand, it requires a careful investigation of the statistical properties of the various regression outputs.

These properties will be derived in a general framework in which the latent model is not necessarily a regression model.

The paper is organized as follows. In the second section we discuss the properties of simulated series and, in particular, we establish a generalised central limit theorem and a law of large numbers for functions of simulations and observed variables. The estimation problems are considered in section 3 ; in particular the asymptotic properties of the second stage estimators are given and compared with that of the maximum lielihood estimators. In section 4 we introduce the notion of simulated residuals and we explain how the residual plots can be used for detecting specifications errors such as : omitted variables, outliers, heteroscedasticity ; we also compare the practical usefulness of these simulated residuals and that of the generalised residuals introduced by CHESHER-IRISH (1984) and GOURIEROUX-MONFORT-RENAULT-TROGNON (1984-a).

Section 5 is devoted to various test procedures ; in particular it is shown how it is possible to correctly use the score test principle in the second stage. Various technical proofs are gathered in appendices.

2. LATENT MODEL AND SIMULATIONS.

2.a - The model.

Three kinds of variables appear in the model : the exogenous variables, the unobservable (or latent) endogenous variables and the observable endogenous variables. x_i , $i = 1,\ldots,n$ denote the d_o-dimensional vectors of exogenous variables, y_i^* , $i = 1,\ldots,n$, denote the d_1-dimensional vectors of latent endogenous variables and y_i , $i = 1,\ldots,n$ are the d_2-dimensional vectors of observable endogenous variables.

It is assumed that the vectors $(y_i^{*\prime},x_i^{*\prime})'$ are independently and identically distributed ; no assumption is made on the true marginal distribution of x_i , but the conditional distribution of y_i^* given x_i is assumed to belong to a family whose p.d.f. are $l^*(y^*/x;\theta)$, $\theta \in \Theta \subset R^K$.

The endogenous variables are defined by :

$$(2.1) \quad y_i = g(y_i^*)$$

The previous assumptions imply that the $(y_i',x_i')'$, $i = 1,\ldots,n$ are independently and identically distributed ; the conditional distribution of y_i given x_i belongs to a family whose p.d.f. are denoted by $l(y/x;\theta)$. It is also assumed that θ is identifiable from this conditional distribution. Finally, it is assumed that there exists a consistent, asymptotically normal estimator of θ , denoted by $\hat{\theta}_n$, which is obtained by maximising a differentiable objective function of the following type :

$$(2.2) \quad \sum_{i=1}^{n} k(y_i,x_i,\theta)$$

More precisely, it is assumed that $\hat{\theta}_n$ is a solution of :

$$\sum_{i=1}^{n} \frac{\partial k}{\partial \theta}(y_i,x_i,\theta) = 0 ,$$

and that :

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = J^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial k}{\partial \theta}(y_i, x_i, \theta) + o_p(1)$$

where $J = E_{\theta_o} - \frac{\partial^2 k}{\partial \theta \, \partial \theta'} (y, x, \theta_o)$ and θ_o is the true value of θ .

An important class of models satisfying the previous assumptions is that in which the latent model is an univariate or multivariate gaussian linear model ;

with such a latent model and with a suitable g function it is possible to reach many usual models such as : univariate or multivariate probit models, simple or generalised tobit models, one-market or multi-market disequilibrium models....

2.b - A generalised central limit theorem.

The statistical procedures proposed in the following sections are based on simulated variables z_{in} , $i = 1,\dots,n$ independently drawn from the conditional distributions $l(y_i^*/y_i, x_i; \hat{\theta}_n)$, i.e. from the conditional distributions of the latent endogenous variables, in which the parameter is replaced by the value of the estimator $\hat{\theta}_n$ introduced in the previous section.

The asymptotic properties of these statistical procedures rest upon the asymptotic behaviour of random variables of the form :

$$(2.3) \quad \frac{1}{n} \sum_{i=1}^{n} h(z_{in}, y_i, x_i)$$

where h is a H-dimensional function.

If the z_{in} , $i = 1,\dots,n$, were drawn from the true conditional distribution $l(y^*_i/y_i,x_i,\theta_o)$, the usual central limit theorem would apply ; but, since the estimator $\hat{\theta}_n$ used in the simulations of the z_{in}'s is a random variable, depending on (y_i,x_i) , $i = 1,\dots,n$, the z_{in}'s are correlated and it is necessary to establish a generalisation of the classical central limit theorem.

THEOREM 2.4 (generalised central limit theorem)

Let ξ_n be the random vector defined by :

$$\xi_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [h(z_{in},y_i,x_i) - E_{\theta_o} h(y^*,y,x)]$$

where E_{θ_o} is the expectation operator with respect to the true distribution of (y^*,x) .

Under regularity assumptions given in appendix 1, ξ_n converges in distribution, as $n \to \infty$, to the zero-mean normal distribution, whose covariance matrix is :

$$V_{\theta_o} h - V_{\theta_o} E_{\theta_o}(h/y,x)$$

$$+ V_{\theta_o}\left\{E_{\theta_o}(h/y,x) + E_{\theta_o}\left[\frac{\partial}{\partial\theta'} E_{\theta_o}(h/y,x)\right] J^{-1} \frac{\partial k}{\partial\theta}(y,x,\theta_o)\right\}$$

Proof : See appendix 1 □

It is straightforward to verify that this theorem contains, as a special case, the usual central limit theorem if $\hat{\theta}_n = \theta_o$; indeed, in this case, we can choose, for instance $k = - \|\theta - \theta_o\|^2$ which implies $\frac{\partial k}{\partial\theta}(y,x,\theta_o) = 0$ and, therefore, the covariance matrix of ξ_n becomes :

$$V_{\theta_o} h - V_{\theta_o} E_{\theta_o}(h/y,x) + V_{\theta_o} E_{\theta_o}(h/y,x) = V_{\theta_o} h$$

Corollary 2.5 :

Under regularity assumptions given in appendix 1, the asymptotic covariance matrix of ξ_n can be written :

$$V_{\theta_o} h - V_{\theta_o} E_{\theta_o}(h/y,x) + V_{\theta_o}\left\{E_{\theta_o}(h/y,x) + E_{\theta_o}\left[h.\frac{\partial \text{ Log } l(y^*/y,x,\theta_o)}{\partial\theta'}\right] J^{-1} \frac{\partial k}{\partial\theta}(y,x,\theta_o)\right\}$$

Proof :See appendix 1.

Corollary 2.6 (weak law of large numbers)

Under the same assumptions as in theorem 2.4 ,

$\frac{1}{n}\sum_{i=1}^{n} h(z_{in},y_i,x_i)$ converges in probability to $E_{\theta_o} h(y^*,y,x)$ as $n \to \infty$

Proof : $\frac{1}{n}\sum_{i=1}^{n}\left[h(z_{in},y_i,x_i) - E_{\theta_o} h(y^*,y,x)\right] = \frac{1}{\sqrt{n}}\xi_n$ and the result follows from theorem 2.4 □

Note that under a modified set of assumptions it is also possible to establish a strong law of large numbers (see Appendix 2).

3. ESTIMATION.

3.a - Theoretical results.

As mentioned in the previous section, θ can be estimated by $\hat{\theta}_n$ and a priori, there is no reason to consider another estimator, specially if $\hat{\theta}_n$ is asymptotically efficient. However, if we want to use mechanically the test procedures available for the latent model when y_i^* is replaced by z_{in}, we are implicitly led to consider estimators of θ based on the z_{in}'s, $i = 1,\dots,n$. For instance, if the latent model is a linear model $y_i^* = x_i b + u_i$, we shall have to consider the least squares estimators of b obtained from a regression of the vector z_{in}, $i = 1,\dots,n$, on the exogenous variables. More generally, it is necessary to study the properties of the estimator $\tilde{\theta}_n$ obtained by maximising the likelihood function of the latent model in which the y_i^*, $i = 1,\dots,n$ have been replaced by the z_{in}, $i = 1,\dots,n$.

In the sequel we assume that the z_{in}, $i = 1,\dots,n$ have been independently drawn from the distributions $l(y_i^*/y_i, x_i, \hat{\theta}_n)$, where $\hat{\theta}_n$ is the maximum likelihood estimator of θ, in the observable model. In other words, we assume that :

$$(3.1) \quad k(y,x,\theta) = \text{Log } l(y/x;\theta) .$$

It follows that $J = E_{\theta_0}\left(- \dfrac{\partial^2 \text{Log } l}{\partial\theta\,\partial\theta'} \right)$ is the Fisher information matrix in the observable model. The estimator $\tilde{\theta}_n$ that we are going to study is obtained by maximizing :

$$(3.2) \quad \sum_{i=1}^{n} \text{Log } l^*(z_{in}/x_i;\theta)$$

THEOREM 3.3 :

> Under regularity assumptions given in appendix 2, $\tilde{\theta}_n$ is a strongly consistent estimator of θ_o .

Proof : See appendix 2.

Once the consistency is established the asymptotic normality is a consequence of the generalised central limit theorem 2.4.

THEOREM 3.4 :

> Under regularity conditions given in appendices 1 and 2, $\sqrt{n}\,(\tilde{\theta}_n - \theta_o)$ is asymptotically normally distributed ; the limit normal distribution is zero mean, its covariance matrix is :
>
> $$\sum = I^{-1} - I^{-1} J\, I^{-1} + J^{-1}$$
>
> where I is the Fisher information matrix in the latent model and J is the Fisher information matrix in the observable model ; both these matrices are evaluated at θ_o .

Proof :

$\tilde{\theta}_n$ is solution of

$$\sum_{i=1}^{n} \frac{\partial \text{ Log } l^*(z_{in}/x_i;\theta)}{\partial \theta} = 0 \quad .$$

From an expansion around θ_o we obtain :

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \operatorname{Log} l^*(z_{in}/x_i;\theta_o)}{\partial\theta} + \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \operatorname{Log} l^*}{\partial\theta\,\partial\theta'} \sqrt{n}\,(\tilde{\theta}_n - \theta_o) = o_p(1)$$

or

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \operatorname{Log} l^*(z_{in}/x_i;\theta_o)}{\partial\theta} - I \sqrt{n}\,(\tilde{\theta}_n - \theta_o) = o_p(1) .$$

$$\sqrt{n}\,(\tilde{\theta}_n - \theta_o) = \frac{1}{\sqrt{n}} I^{-1} \sum_{i=1}^{n} \frac{\partial \operatorname{Log} l^*(z_{in}/x_i;\theta_o)}{\partial\theta} + o_p(1) .$$

The asymptotic normality is a consequence of theorem 2.4 ; the asymptotic mean is zero and the asymptotic covariance matrix is obtained from the general formula of corollary 2.6, by replacing h by $I^{-1} \dfrac{\partial \operatorname{Log} l^*(y^*/x;\theta_o)}{\partial\theta}$.
First we have to compute

$$E_{\theta_o}(h/y,x) = I^{-1} E_{\theta_o}\left[\frac{\partial \operatorname{Log} l^*(y^*/x;\theta_o)}{\partial\theta} \Big/ y,x\right]$$

$$= I^{-1} \frac{\partial \operatorname{Log} l(y/x;\theta_o)}{\partial\theta}$$

(See e.g. MONFORT (1982) p. 73).

Thus, the asymptotic covariance of $\sqrt{n}\,(\tilde{\theta}_n - \theta_o)$ is :

$$\sum = V_{\theta_o} h - V_{\theta_o} E_{\theta_o}(h/y,x) + V_{\theta_o}\left[(I^{-1} + AJ^{-1}) \frac{\partial \operatorname{Log} l(y/x;\theta_o)}{\partial\theta}\right] ,$$

where $A = E_{\theta_o}\left[h \cdot \frac{\partial \text{ Log } l(y^*/y,x;\theta_o)}{\partial\theta}\right]$.

Therefore, we get

$$\sum = I^{-1} - I^{-1}JI^{-1} + (I^{-1}+AJ^{-1})J(I^{-1}+J^{-1}A') \quad .$$

Let us now compute A .

$$A = I^{-1} E_{\theta_o}\left[\frac{\partial \text{ Log } l^*(y^*/x;\theta_o)}{\partial\theta} \cdot \frac{\partial \text{ Log } l(y^*/y,x;\theta_o)}{\partial\theta'}\right]$$

$$= I^{-1} E_{\theta_o}\left[\frac{\partial \text{ Log } l^*(y^*/x;\theta_o)}{\partial\theta} \cdot \frac{\partial \text{ Log } l^*(y^*/x;\theta_o)}{\partial\theta'}\right]$$

$$- I^{-1} E_{\theta_o}\left[\frac{\partial \text{ Log } l^*(y^*/x;\theta_o)}{\partial\theta} \cdot \frac{\partial \text{ Log } l(y/x;\theta_o)}{\partial\theta'}\right].$$

The first expectation is equal to I. Taking the conditional expectation given y and x, it is easily seen that the second expectation is equal to J .
Therefore :

$$A = I^{-1}(I - J) \quad ,$$

and

$$A J^{-1} = I^{-1}(I - J) J^{-1}$$
$$= J^{-1} - I^{-1} \quad , \qquad \square$$

As expected, $\tilde{\theta}_n$ is asymptotically less efficient than $\hat{\theta}_n$, since the asymptotic covariance matrix of $\hat{\theta}_n$ is J^{-1} and $\Sigma - J^{-1} = I^{-1}(I-J)I^{-1}$ is positive. Moreover it is clear that $\tilde{\theta}_n$ is asymptotically efficient if, and only if, $I = J$, that is if and only if, g is a sufficient statistic for θ [see e.g MONFORT (1982), p. 74].

In order to have a more precise insight of the efficiency loss when using $\tilde{\theta}_n$ instead of $\hat{\theta}_n$, let us consider the one dimension case. If θ is a scalar parameter the formula giving Σ can be written :

$$(3.5) \quad V_{as}\,\tilde{\theta}_n = V_{as}\,\hat{\hat{\theta}}_n - \frac{\left(V_{as}\,\hat{\hat{\theta}}_n\right)^2}{V_{as}\,\hat{\theta}_n} + V_{as}\,\hat{\theta}_n$$

where $\hat{\hat{\theta}}_n$ is the maximum likelihood estimator of θ in the latent model (which is not computable since the y_i^* s are unobserved) and where V_{as} means asymptotic variance.

Let us denote by r the ratio

$$(3.6) \quad r = \frac{V_{as}\,\hat{\hat{\theta}}_n}{V_{as}\,\hat{\theta}_n} \qquad 0 < r < 1$$

r is the asymptotic relative efficiency of $\hat{\theta}_n$ with respect to $\hat{\hat{\theta}}_n$, the efficiency loss being a consequence of the unobservability of y_i^* .
Let us define ϱ by :

$$(3.7) \quad \varrho = \frac{V_{as}\,\hat{\theta}_n}{V_{as}\,\tilde{\theta}_n} \qquad 0 < \varrho < 1 \quad .$$

ϱ is the asymptotic relative efficiency of $\tilde{\theta}_n$ with respect to $\hat{\theta}_n$, i.e. to the best computable estimator.

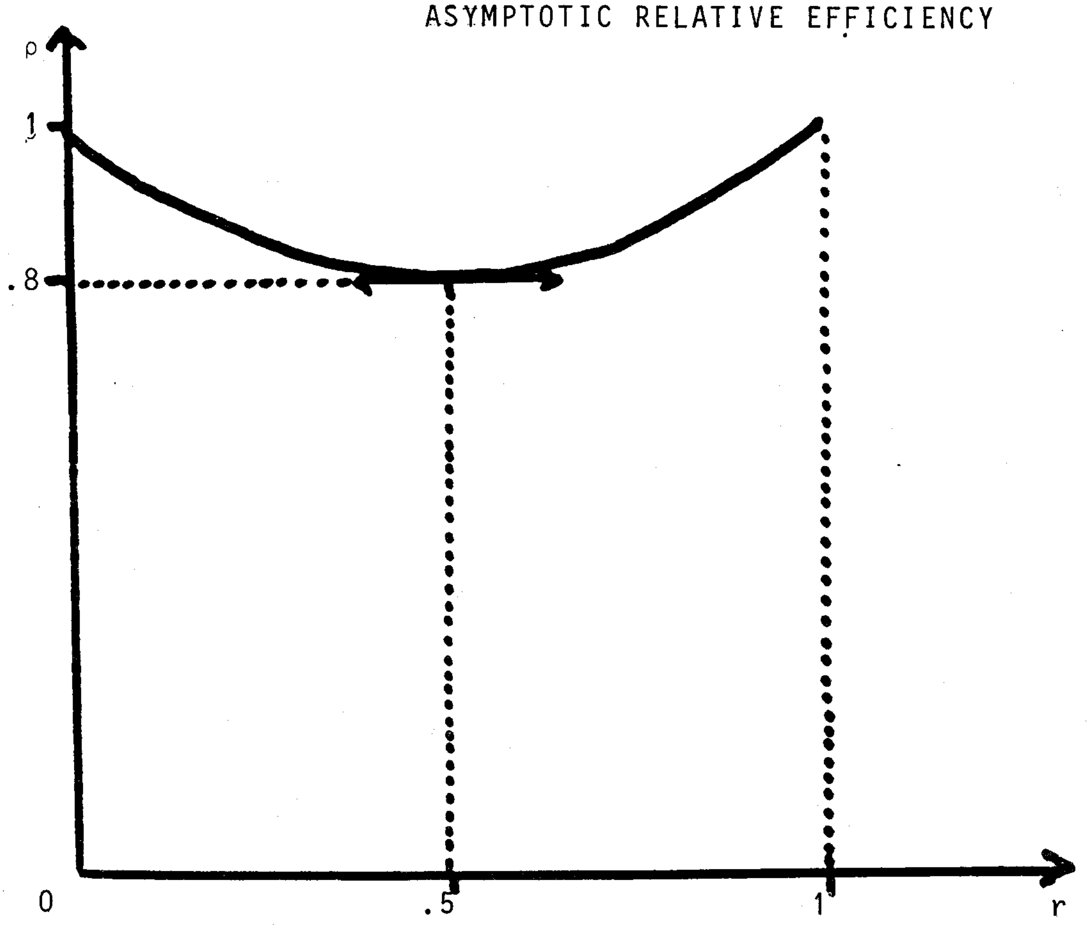
From (3.5) we obtain :

$$(3.8) \quad \varrho = \frac{1}{1 + r - r^2} \quad .$$

In other words the asymptotic relative efficiency ρ of $\tilde{\theta}_n$, due to the simulation procedure, is a function of the asymptotic relative efficiency r , due to the unobservability of the latent model.

Moreover function (3.8) is readily seen to be symmetric with respect to $r = .5$; it is equal to 1 for $r = 0$ and $r = 1$ and its minimum, reached for $r = .5$, is equal to .8 .

FIGURE 2.



The maximal efficiency loss is equal to 20%, in terms of variance, i.e. about 10% in terms of standard deviation which does not seem unreasonable. The loss is small if r is near 1 , i.e. if the observable model nearly catches the whole information or, on the contrary, if r is near 0 , i.e. if the observable model nearly looses the whole information.

This is intuitively clear since, if the whole information is catched, $y = g(y^*)$ is a sufficient statistic and, therefore, the conditional distribution of y_i^+ given y_i is known ; on the other hand is the whole information is lost, $y = g(y^*)$ is and there is no point, anyway, in basing inference on y .

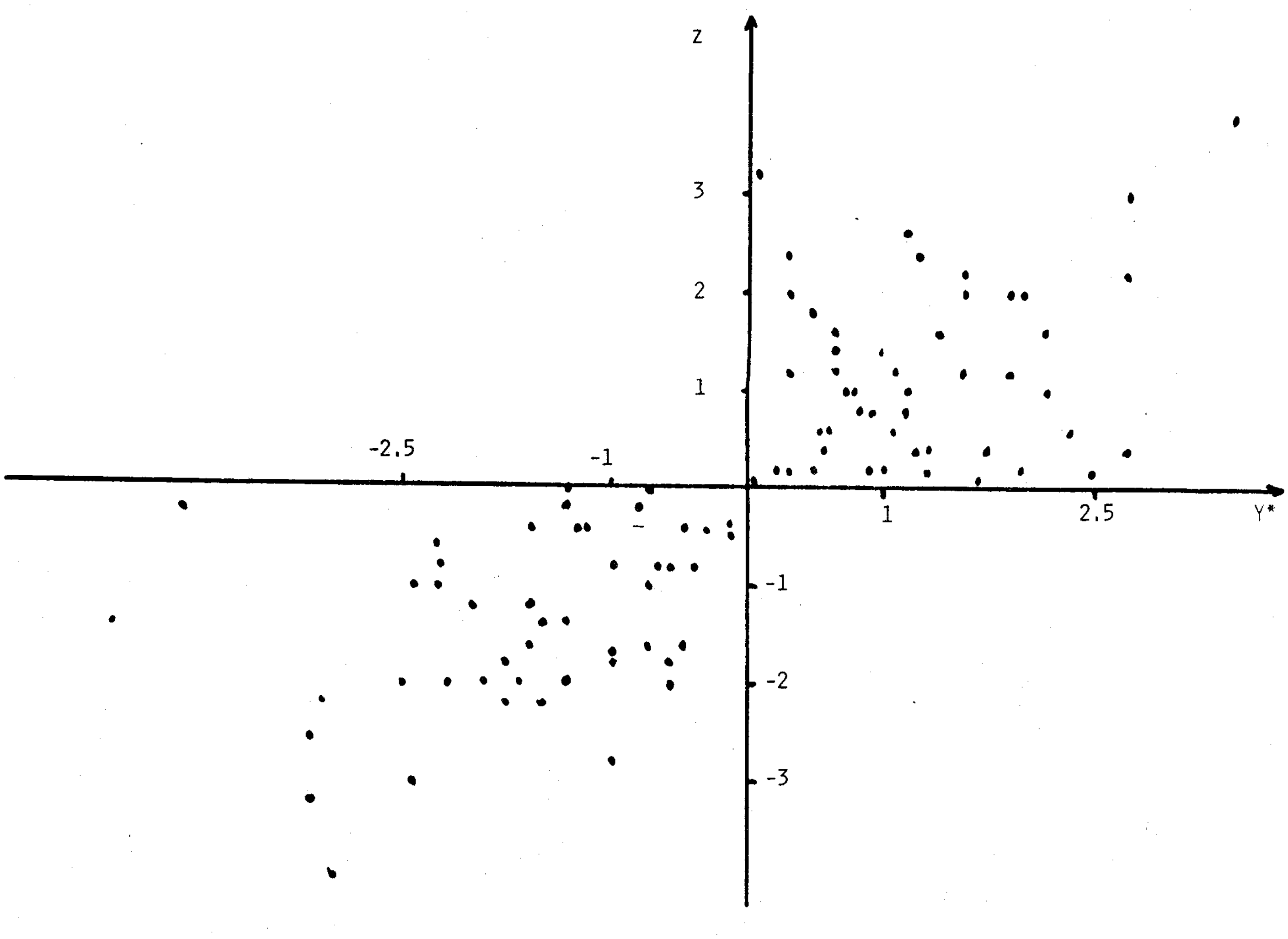
3.b - An illustration.

The artificial probit model presented in section 1 is used in this subsection as an empirical illustration of the previous theoretical results.

Except for their signs, the simulated latent variables z_{in} are not very close to the latent variables y_i^* as it is shown in the following figure :

FIGURE 3

COMPARISON OF LATENT AND SIMULATED LATENT VARIABLES



Nevertheless the maximum likelihood estimator $\hat{\theta}' = (\hat{\alpha},\hat{\beta})$ does not differ much from $\tilde{\theta}' = (\tilde{\alpha},\tilde{\beta})$. We have seen in the introduction that, on the contrary, the computed variance-covariance matrix of $\tilde{\theta}$ given by the least squares package is far from the estimated asymptotic covariance matrix of $\hat{\theta}$. For the present simulation, we have obtained :

$$\tilde{V}(\tilde{\theta}) = \tilde{I}^{-1} = \begin{pmatrix} .0134 & .0045 \\ .0045 & .0045 \end{pmatrix}$$

$$\hat{V}(\hat{\theta}) = \hat{J}^{-1} = \begin{pmatrix} .0373 & .0163 \\ .0163 & .0223 \end{pmatrix}$$

According to theorem 3.4 and to corollary 2.6 a consistent estimate of the asymptotic covariance matrix of $\tilde{\theta}$ is :

$$V^*(\tilde{\theta}) = \tilde{I}^{-1} - \tilde{I}^{-1}\hat{J}\,\tilde{I}^{-1} + \hat{J}^{-1} = \begin{pmatrix} .0458 & .0194 \\ .0194 & .0258 \end{pmatrix}$$

Thus the computed standard errors of $\tilde{\alpha}$ and $\tilde{\beta}$ deduced from $\tilde{V}(\tilde{\theta})$ are respectively (.119) and (.068) whereas from $V^*(\tilde{\theta})$ they are (.215) and (.161). The t-ratio for the hypothesis $H_0 = (\beta = 1)$ computed from $\tilde{V}$ is -2.044 while the asymptotic t-ratio deduced from V^* is in fact - .863. If the test of size .05 is based on the critical value of the standard normal distribution (- 1.96) the wrong regression results reject H_0 whereas the correct asymptotic test does not. This empirical result shows that one should be cautious about the standard use of the tests that are produced by the computer in the second step of the simulation procedure.

The previous data allow to compute an evaluation of the asymptotic relative efficiency of $\tilde{\theta}$ with respect to $\hat{\theta}$. For α and β we obtain respectively :

$$\rho_\alpha = .0373/.0458 = 81\ \%$$

$$\rho_\beta = .0223/.0258 = 86\ \%$$

4 - SIMULATED RESIDUAL PLOTS.

In this section we restrict the analysis to the case where the latent process is a linear regression model :

$$y_i^* = x_i b + u_i$$

Conditionally upon x_i the disturbance u_i is assumed to be normally distributed with zero mean.

In the more specific case where the transformation $g(.)$ is the identity mapping all the latent variables are observable. If, furthermore, the conditional variance of u_i is σ^2, constant over the sample, the ML estimator of b is the OLS estimator. The residual $\hat{u}_i = y_i - x_i \hat{b} = y_i^* - x_i \hat{b}$ is defined as the deviation of y_i (or y_i^*) from the ML estimation of the conditional expectation of y_i given x_i. Since $\hat{b}$ is a consistent estimator of b, $x_i \hat{b}$ is close to $x_i b$ in large samples. $\hat{u}_i$ mimics the disturbance u_i and it is natural to judge the correctness of assumptions on u_i by graphical methods based on residuals. It is a common practice to use some residual plots for the detection of, say, outliers, heteroscedasticity, omitted variables, etc....

In the general model where $g(.)$ is not a one to one mapping it is also important to examine the correctness of assumptions about the u_i's. But the difficulty comes from the fact that y_i^* is not always available and the previous residuals cannot be computed.

Nevertheless, following the procedure adopted in the previous section, some simulated latent variables z_{in} can be drawn from the conditional distribution of the latent endogenous variables given the exogenous and observable endogenous variables in which the parameter is replaced by the value of the ML estimator $\hat{\theta}_n = (\hat{b}_n, \hat{\sigma}_n^2)$ of $\theta = (b, \sigma^2)$. For the present model $(\tilde{b}_n, \tilde{\sigma}_n^2)$ is then obtained by regressing z_{in} on x_i $i = 1,\dots,n$. Let us define the simulated residuals as :

$$\tilde{u}_{in} = z_{in} - x_i \tilde{b}_n \qquad i = 1,\dots,n$$

Since z_{in} has the same asymptotic distribution as y_i^* and $\tilde{b}_n$ is consistent, $\tilde{u}_{in}$ has the same distribution as u_i in large sample. It is then possible to use the simulated residuals as we use the classical ones to detect some underlying features of the residuals.

To show that this can actually be done in practice, we use some very simple examples of Probit and Tobit models with outliers, heteroscedasticity and omitted variables misspecifications.

4.a - Detecting outliers by simulated residual plots.

To begin with, consider the case where the latent data generating process (DGP) is affected by outliers. The latent variables are generated as described in section 1 . The 33th and 66th observations of this DGP have been replaced by $y_{33}^* = +10$ and $y_{66}^* = -10$ instead of $y_{33}^* = -1.24$ and $y_{66}^* = 1.65$. For the Probit model :

$$\begin{cases} y_i = 1 & \text{if} \quad y_i^* > 0 , \\ y_i = 0 & \text{otherwise} \end{cases}$$

and for the Tobit model :

$$\begin{cases} y_i = y_i^* & \text{if} \quad y_i^* > 0 \\ y_i = 0 & \text{otherwise} \end{cases}$$

The modification of the DGP moves y_{33} from 0 to 1 and y_{66} from 1 to 0 in Probit model and moves y_{33} from 0 to + 10 and y_{66} from 1.65 to 0 in Tobit model. Around $i = 33$ the y_i^* are almost all negative and around $i = 66$ the sign of y^* is not so well defined. Thus it seems a priori that it will be more difficult to detect an outlier at the 66th observation than at the 33th one. This presumption is confirmed by figures 4 and 5 which display the simulated residuals plots function of the exogenous variable for Probit and Tobit models.

FIGURE 4

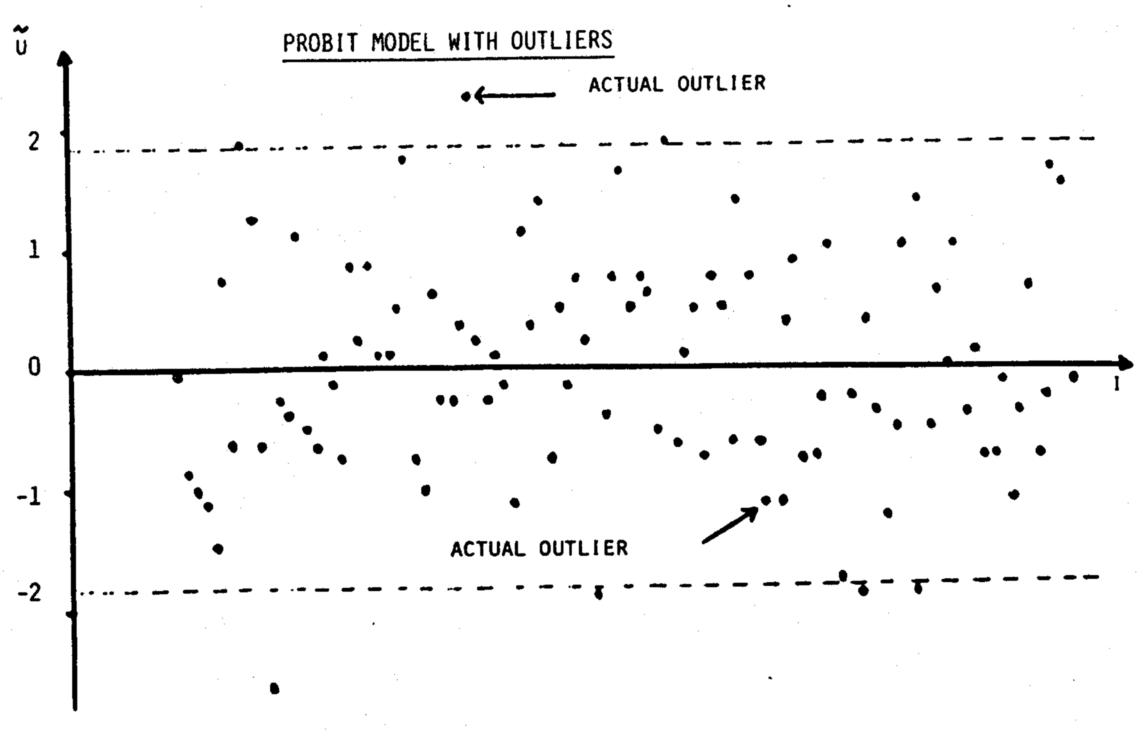
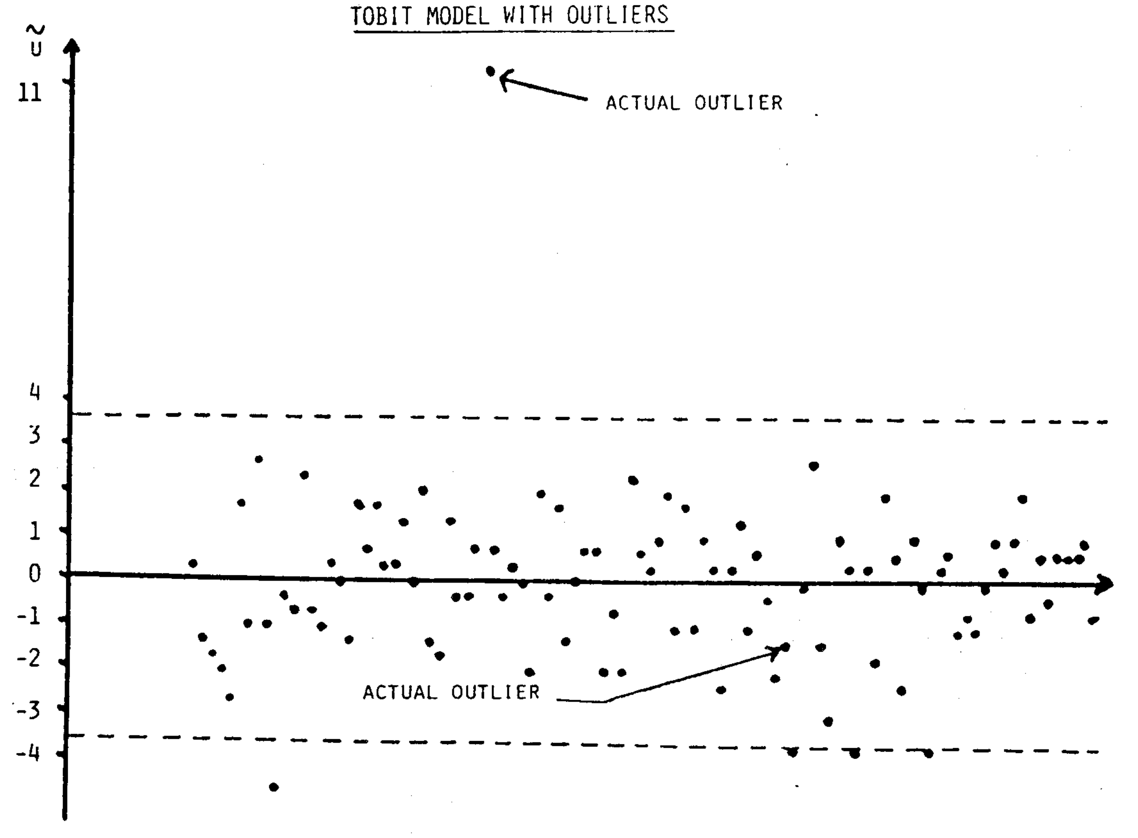


FIGURE 5


TOBIT MODEL WITH OUTLIERS
ũ
11
4
3
2
1
0
-1
-2
-3
-4
ACTUAL OUTLIER
ACTUAL OUTLIER

The dashed horizontal lines delimit a region of width $3.92\,\tilde{\sigma}$, centered at zero, which visualizes the classical confidence region. Out of this region the residuals are presumed to come from an outlier. The 33th observation appears to be out of the model much more clearly in the case of the tobit model than in the case of the probit model even if the estimated standard error of the regression is 1.8 for the former, a value far from the true value equal to 1. This large value of $\tilde{\sigma}$ is due to the outlier.

4.b - Detecting heteroscedasticity by simulated residual plots.

The latent DGP has a conditional variance depending on an exogenous variable :

$$y_i^* = 1 + x_i + x_i u_i \qquad u_i \approx IIN(0,1) .$$

The ML estimator $\hat{\alpha}$ and $\hat{\beta}$ of the intercept α and of the slope parameter β, the simulated latent variables z_{in} and $(\tilde{b},\tilde{\sigma}^2)$ are computed as if the model were homoscedastic.

The plots of the simulated residuals against the exogenous variable are displayed in figures 6 and 7 for Probit and Tobit models. The dashed lines delimit an horizontal cone which is the confidence (asymptotic) region of level 95 % to which the residuals should belong if the disturbances were heteroscedastic proportional to x_i.

FIGURE 6

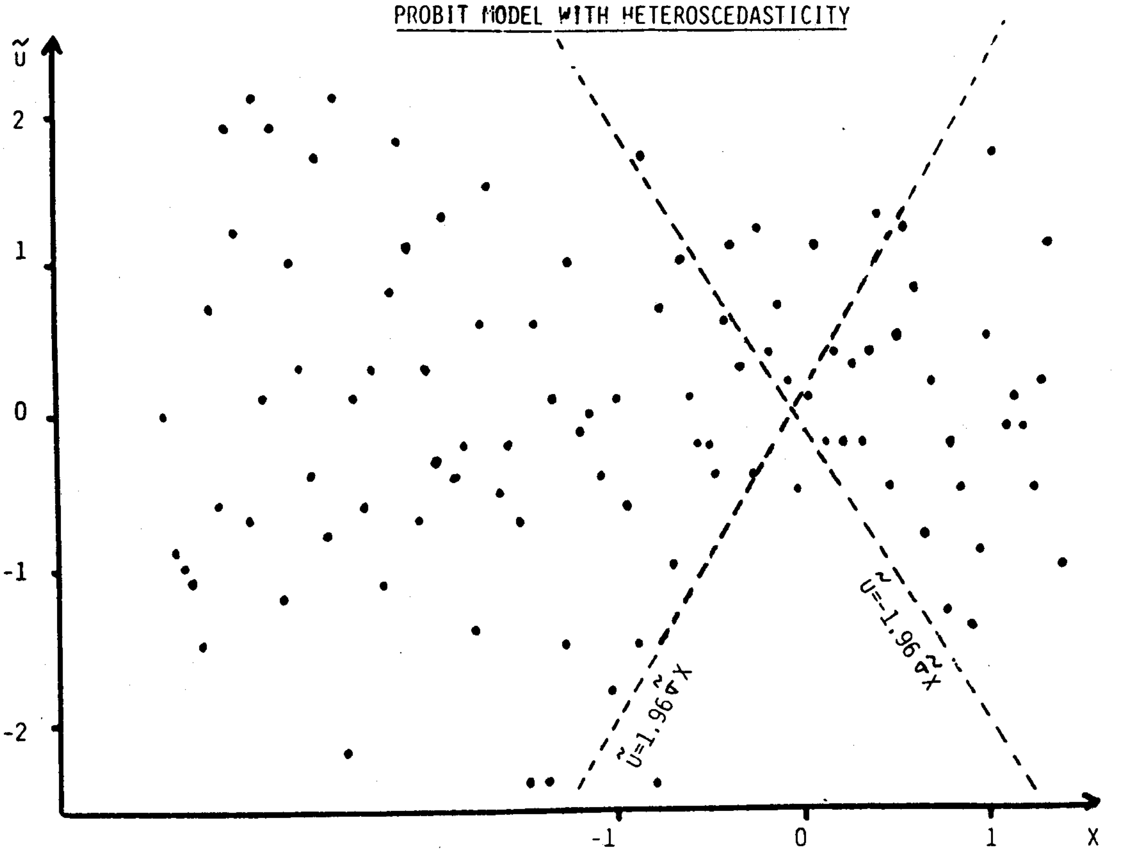
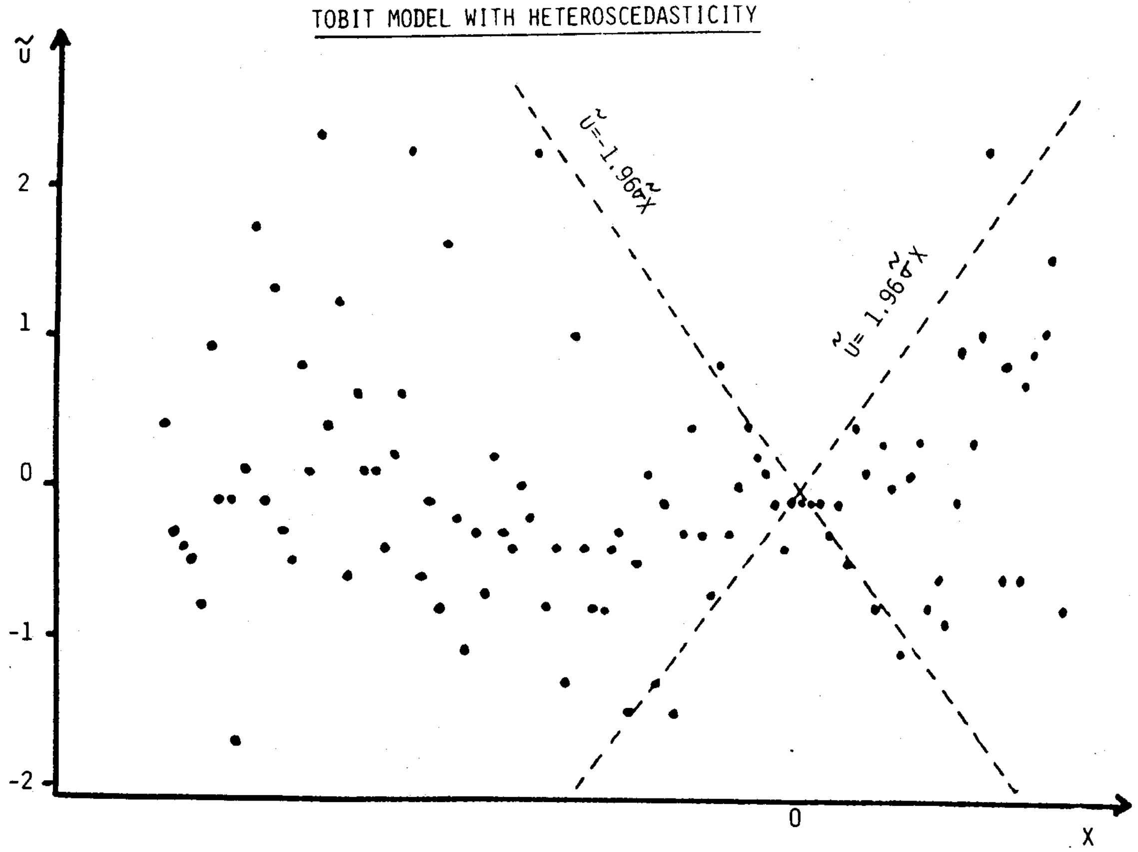

PROBIT MODEL WITH HETEROSCEDASTICITY
ũ
2
1
0
-1
-2
-1
0
1
X
ũ=1.96 σ̃X
ũ=-1.96 σ̃X

FIGURE 7


TOBIT MODEL WITH HETEROSCEDASTICITY
Ũ
2
1
0
-1
-2
Ũ=-1.96σ̃X
Ũ= 1.96σ̃X
0
X

Again, the plots for the tobit model are more conclusive than the ones for the probit model. Nevertheless, even this last one shows some visual clues for possible heteroscedasticity in the latent model. The classical funnel shape appears by careful inspection of those two figures.

4.c - Detecting omitted variables by simulated residual plots.

In this subsection, the latent data generating process has the following form :

$$y_i^* = 1 + x_i + w_i + u_i \quad , \quad u_i \approx \text{I I N } (0,1)$$

with $w_i = \cos(i)$; case(a)

or $w_i = \cos(x_i)$; case(b) .

In case(a) , w_i is almost uncorrelated with x_i (correlation coefficient = .01) ; in case(b), these two exogenous variables are highly correlated (correlation = .85).

The maximum likelihood estimation of the probit and tobit models, the simulated endogenous variables and the simulated residuals are performed as if the variable w_i were not present in the model.

Even when the model is a classical linear model a residual plot is not very informative to detect an omitted variable highly correlated with the exogenous variables kept in the estimated model ; thus case(a) should be more interpretable than case(b). The following figures confirm this is also true for the simulated residual plots. Figures 8 and 9 show a visible positive correlation between the simulated residuals and the omitted variable in case (a). But no real relationship can be detected by figures 10 and 11 in case(b).

FIGURE 8

PLOT OF SIMULATED RESIDUALS FOR A PROBIT MODEL WITH OMITTED VARIABLES ALMOST UNCORRELATED WITH THE MAINTAINED EXOGENOUS VARIABLE

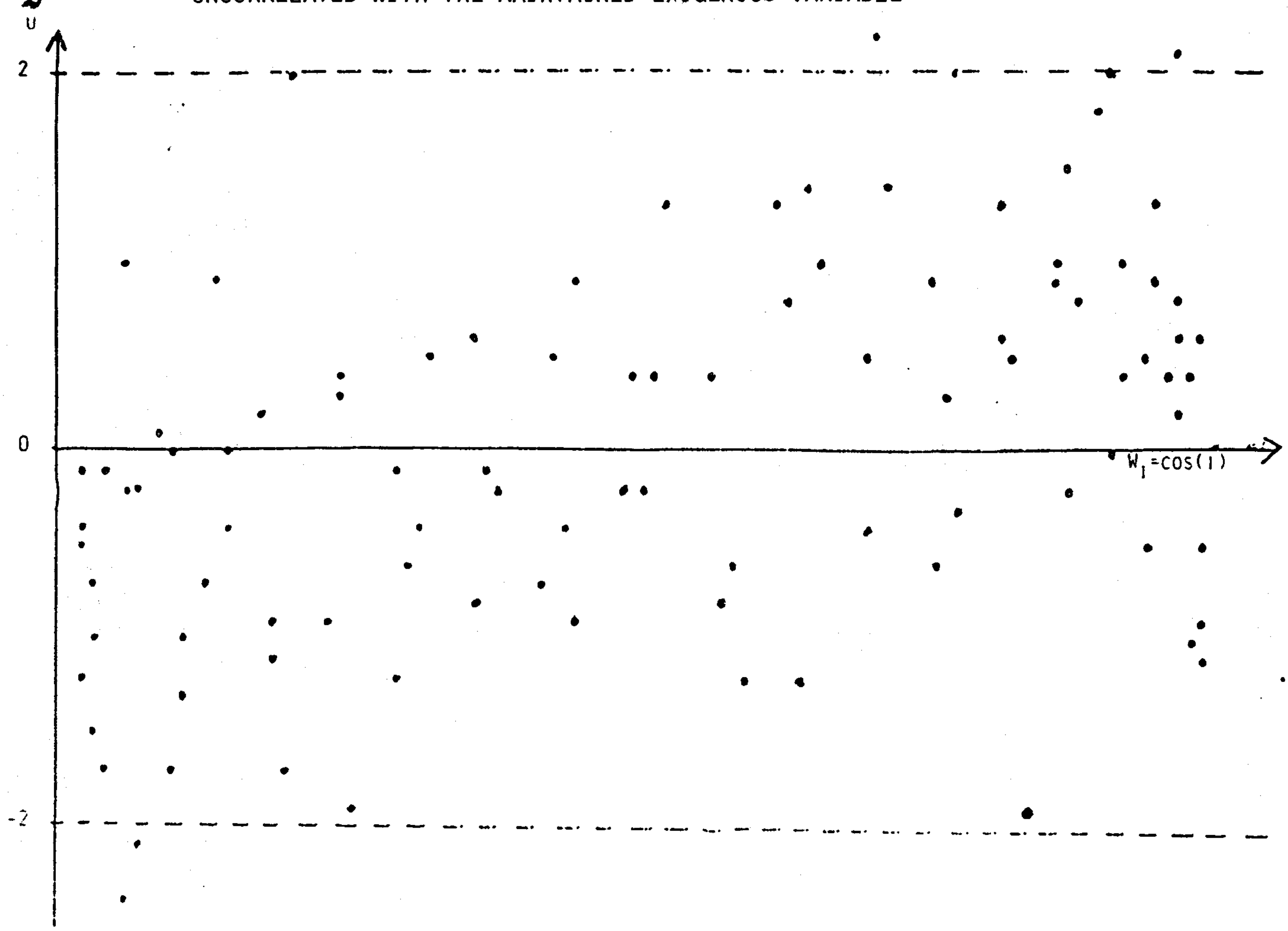


FIGURE 9

PLOT OF SIMULATED RESIDUALS FOR A TOBIT MODEL WITH OMITTED VARIABLES ALMOST UNCORRELATED WITH THE MAINTAINED EXOGENOUS VARIABLE

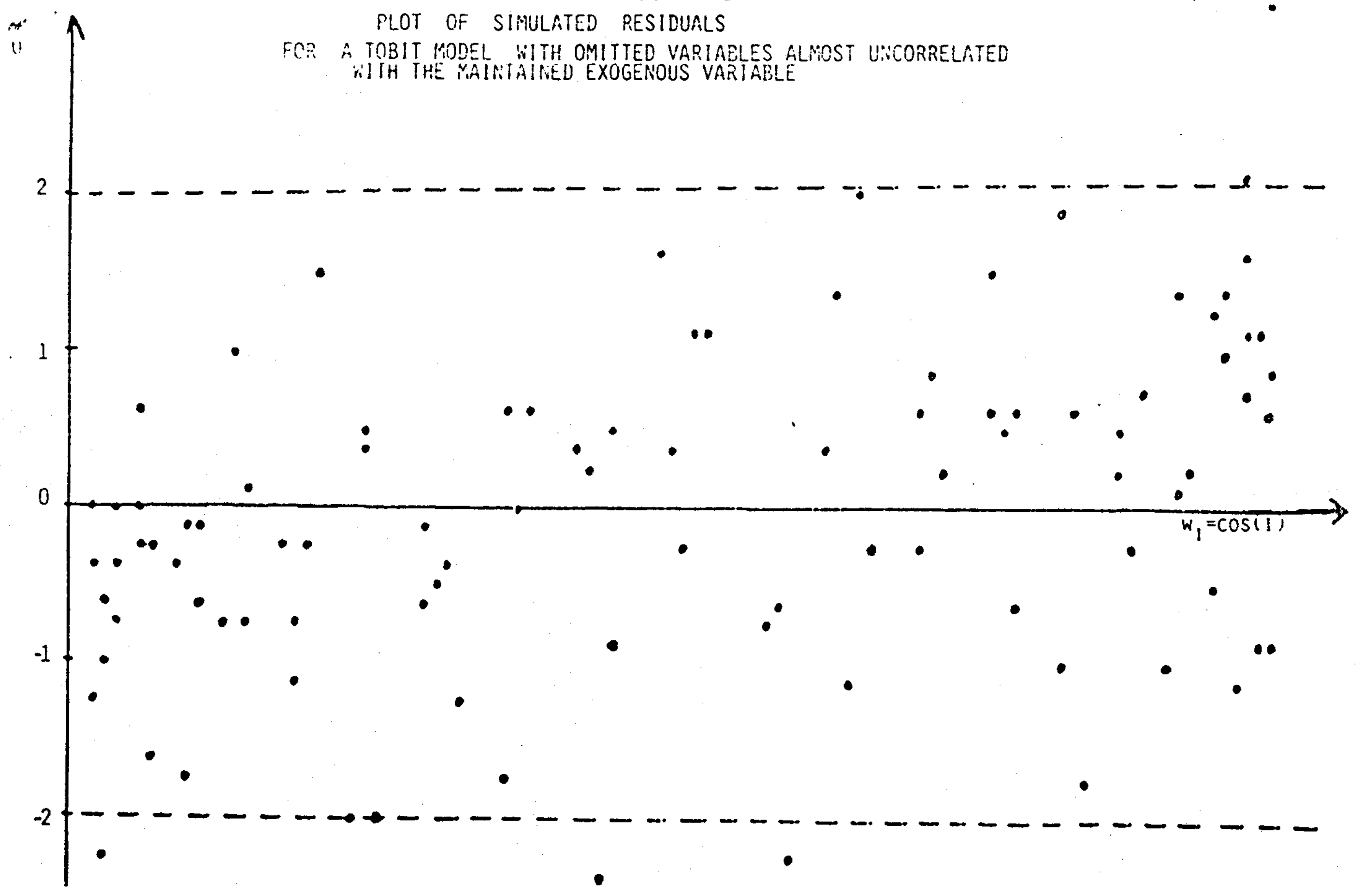


FIGURE 10
PLOT OF SIMULATED RESIDUALS FOR A PROBIT MODEL WITH OMITTED VARIABLES HIGHLY CORRELATED WITH THE MAINTAINED EXOGENOUS VARIABLE

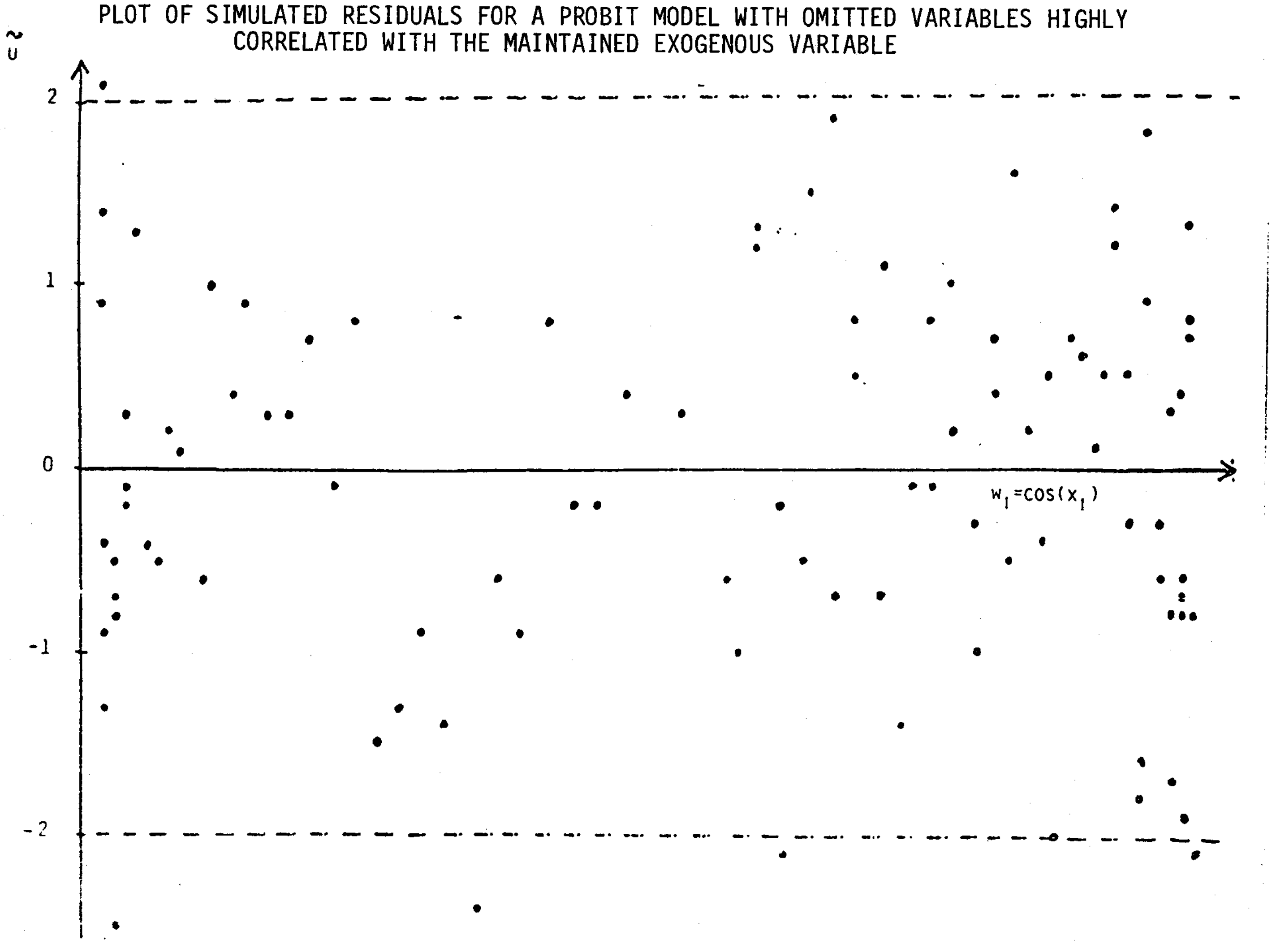
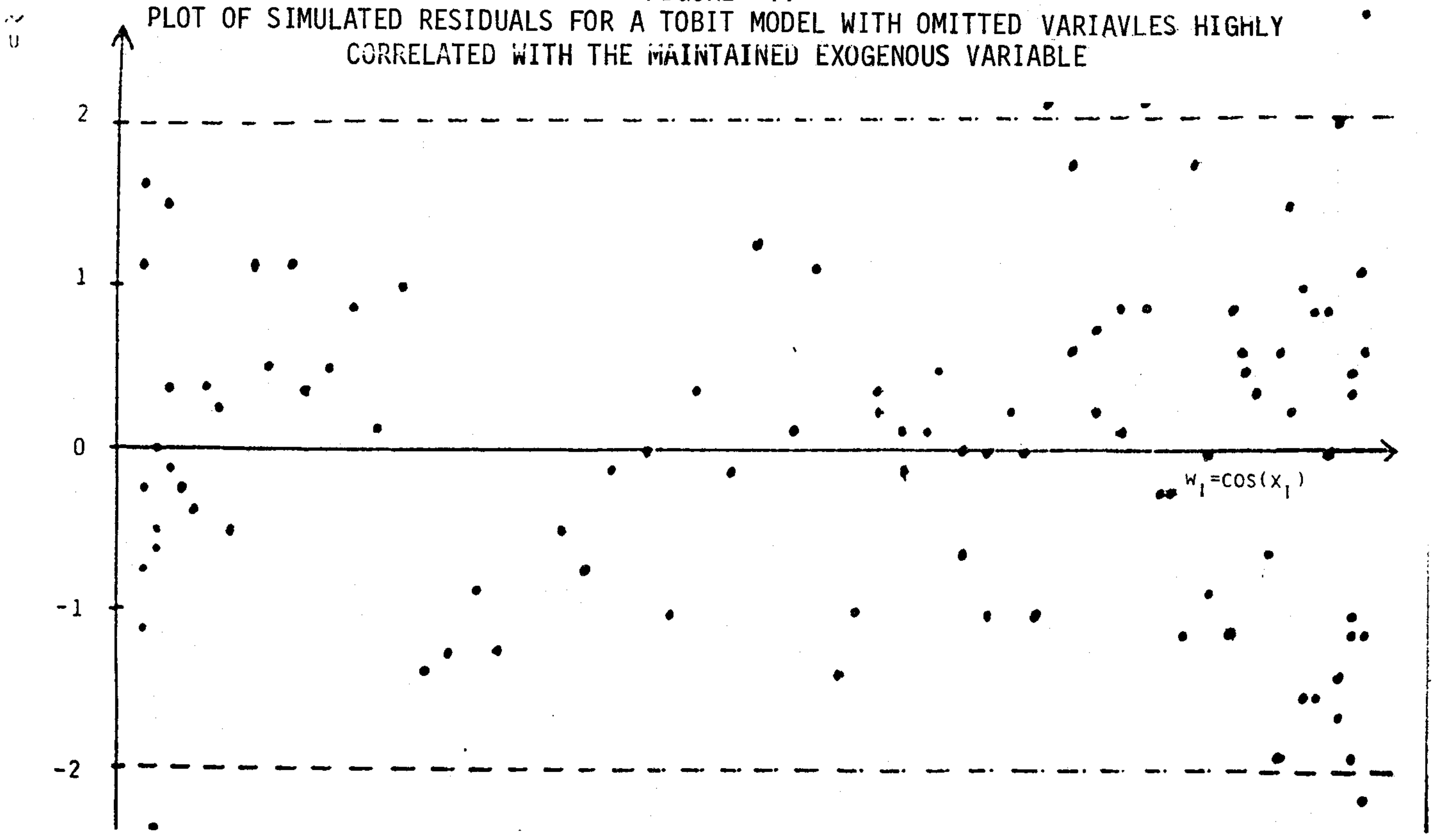


FIGURE 11
PLOT OF SIMULATED RESIDUALS FOR A TOBIT MODEL WITH OMITTED VARIAVLES HIGHLY CORRELATED WITH THE MAINTAINED EXOGENOUS VARIABLE



4.d - Comparison with the generalised residuals.

Recently along the lines initiated by COX-SNELL [1968], CHESHER and IRISH [1984] proposed another procedure which has to be compared with the simulated residuals. They suggested to base empirical diagnostic upon the generalised residuals which are equal to the conditional expectations of the latent disturbances given the observed exogenous and endogenous variables (x_i, y_i) evaluated at the ML estimator $\hat{\theta} = (\hat{b}, \hat{\sigma}^2)$:

$$u_i^* = E_{\hat{\theta}} (u_i / y_i, x_i)$$

For the probit and tobit models the generalised residuals have the following forms :

$$u_i^* = \frac{\varphi(x_i \hat{b})}{\Phi(x_i \hat{b}) [1 - \Phi(x_i \hat{b})]} [y_i - \Phi(x_i \hat{b})]$$

$$u_i^* = (y_i - x_i \hat{b}) \, \mathbb{1}_{(y_i > 0)} - \hat{\sigma} \frac{\varphi\left(\frac{x_i \hat{b}}{\hat{\sigma}}\right)}{1 - \Phi\left(\frac{x_i \hat{b}}{\hat{\sigma}}\right)} \mathbb{1}_{(y_i = 0)}$$

where φ and Φ are respectively the density function and cumulative function of the standard normal.

This last expression shows that for the tobit model and for indexes i such that $y_i > 0$, the generalised residual $u_i^* = y_i - x_i \hat{b}$ and the simulated residuals $\tilde{u}_i = z_{in} - x_i \tilde{b} = y_i - x_i \tilde{b}$ are almost equal, since $\tilde{b}$ is close to $\hat{b}$. On the other hand they differ markedly for the probit or the tobit model when y_i^* is unobservable.

Except in the special case where the latent variable is observed, it has been pointed out that the generalised residuals are difficult to interpret. In order to propose a correct interpretation, the practitioner has to be quite familiarized with the typical patterns of the generalised residual plots.

In the case of Probit model with outliers the corresponding plots do not carry more visual information than the plots of the y_i against x_i. When the disturbances are heteroscedastic, the pattern of the generalised residuals does not change much ; no visual interpretation can be easily obtained. When some variables are omitted CHESHER and IRISH have shown that the generalised residuals have little use if the omitted variables are correlated with the included variables, as it is the case in one of our simulations.

The following figures confirm empirically this fact.

FIGURE 12.

GENERALISED RESIDUAL PLOTS FOR A WELL SPECIFIED PROBIT MODEL

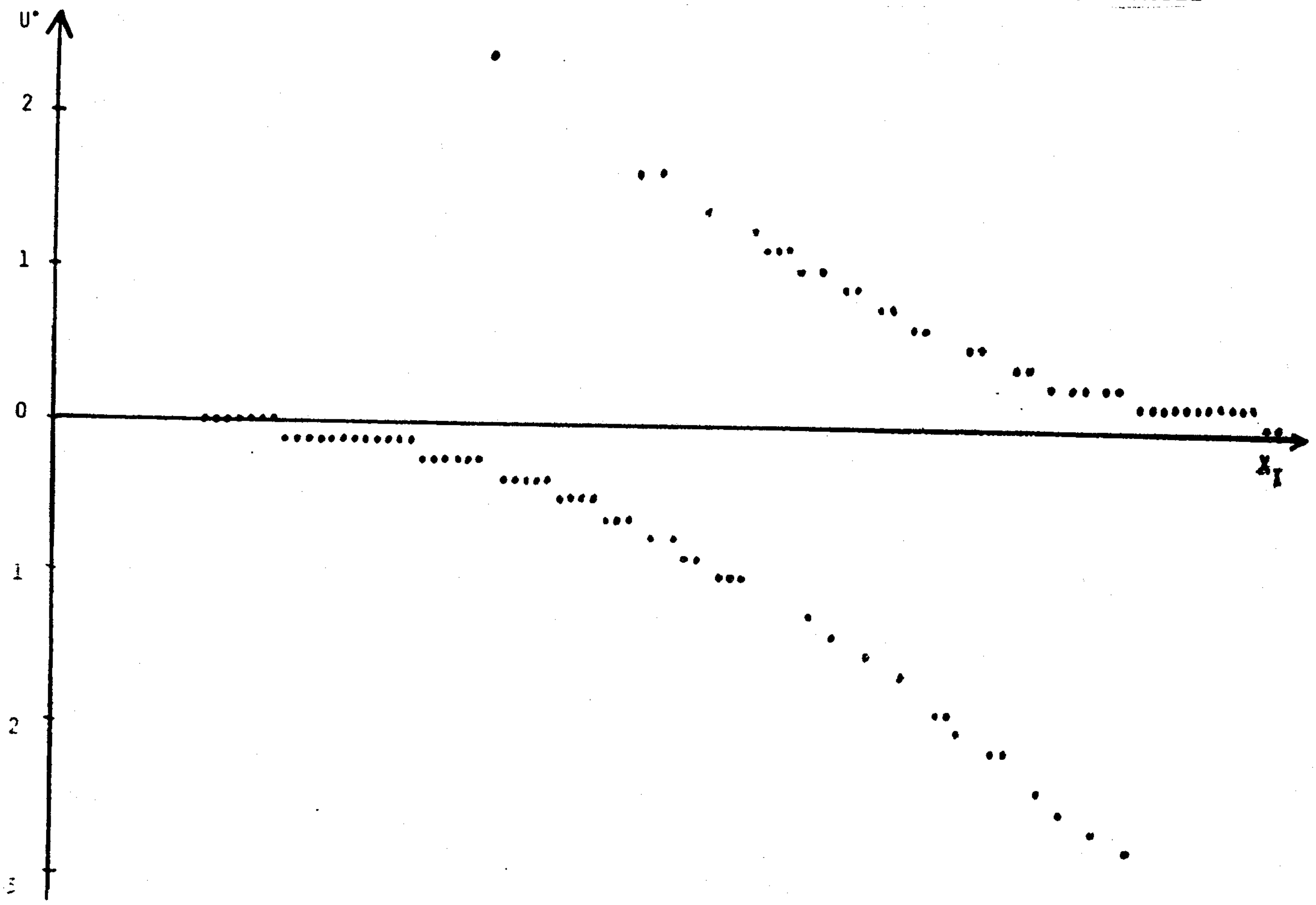


FIGURE 13

GENERALISED RESIDUAL PLOTS FOR A WELL SPECIFIED TOBIT MODEL

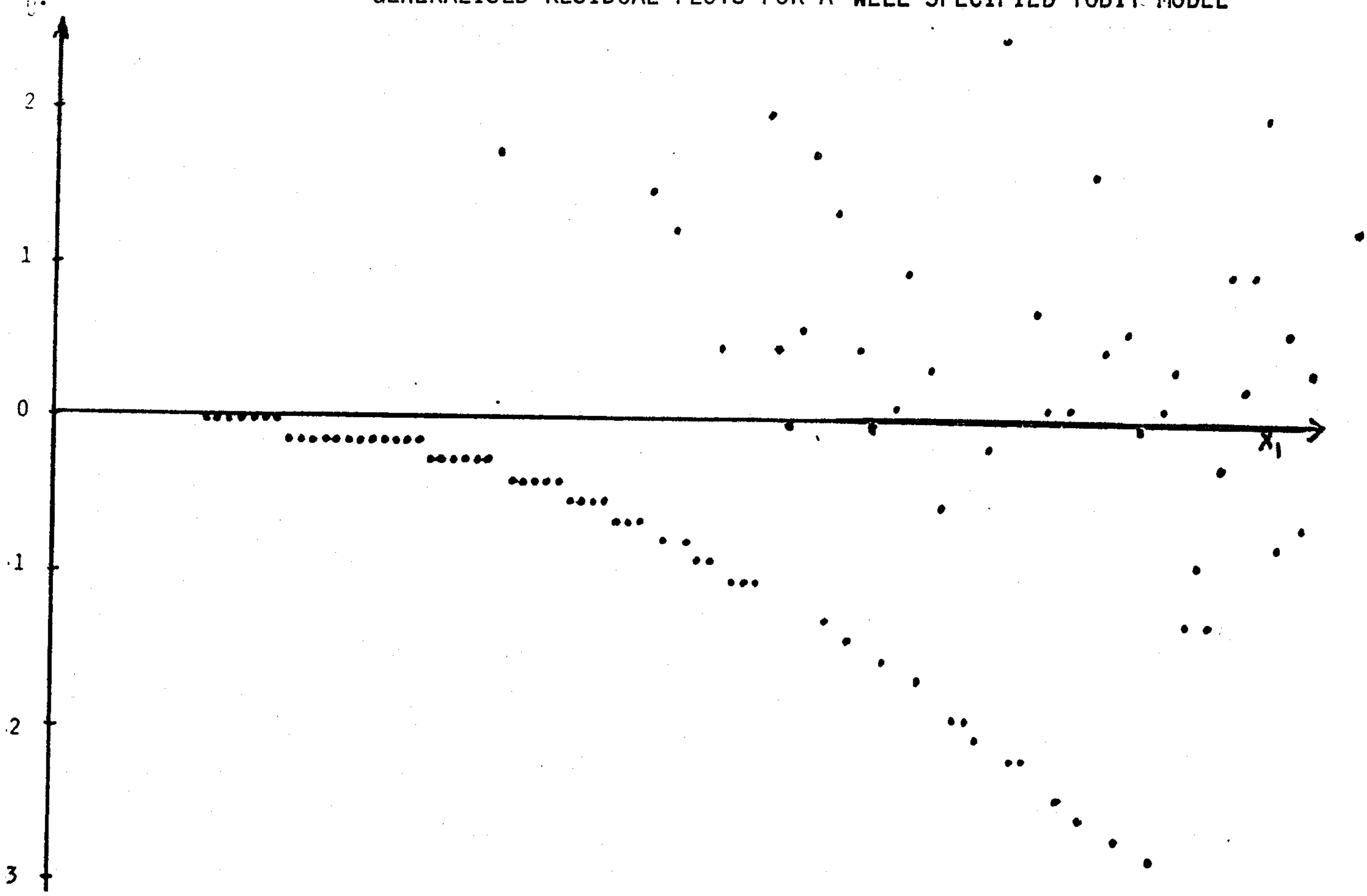


FIGURE 14
GENERALISED RESIDUAL PLOTS FOR A PROBIT MODEL WITH OUTLIERS

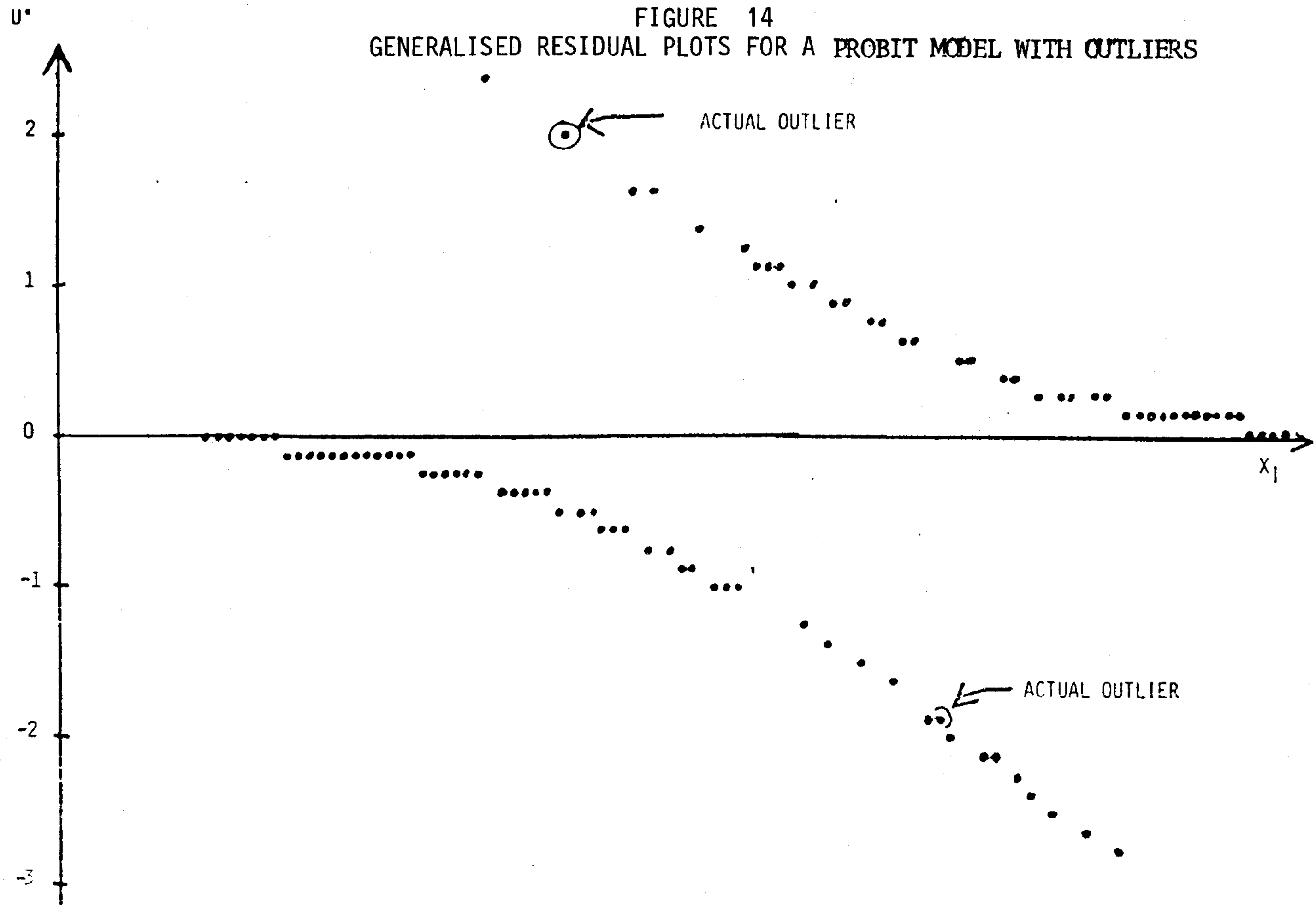


FIGURE 15
GENERALISED RESIDUAL PLOTS FOR A TOBIT MODEL WITH OUTLIERS

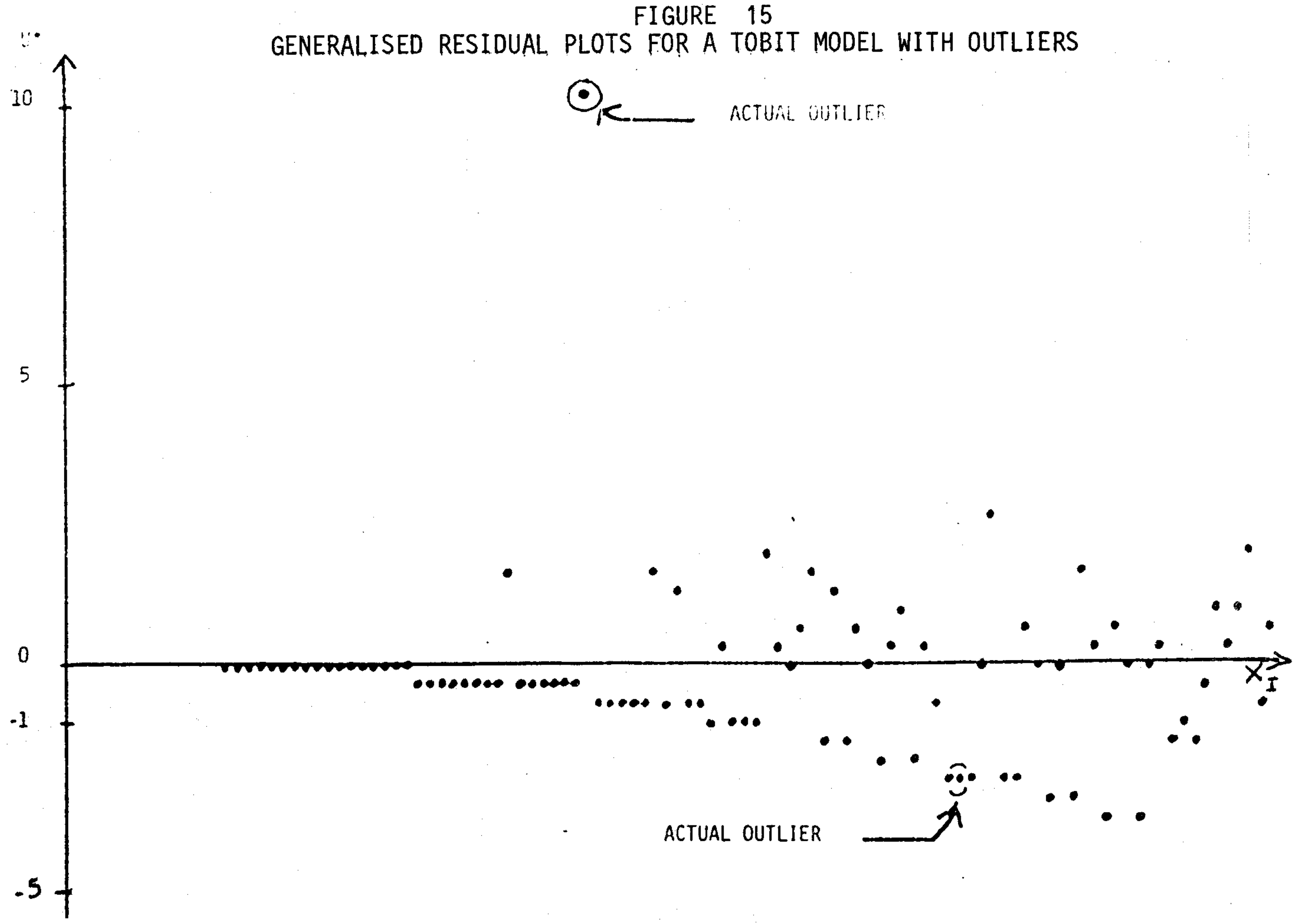


FIGURE 16

GENERALISED RESIDUAL PLOTS FOR A PROBIT MODEL WITH HETEROSCEDASTICITY

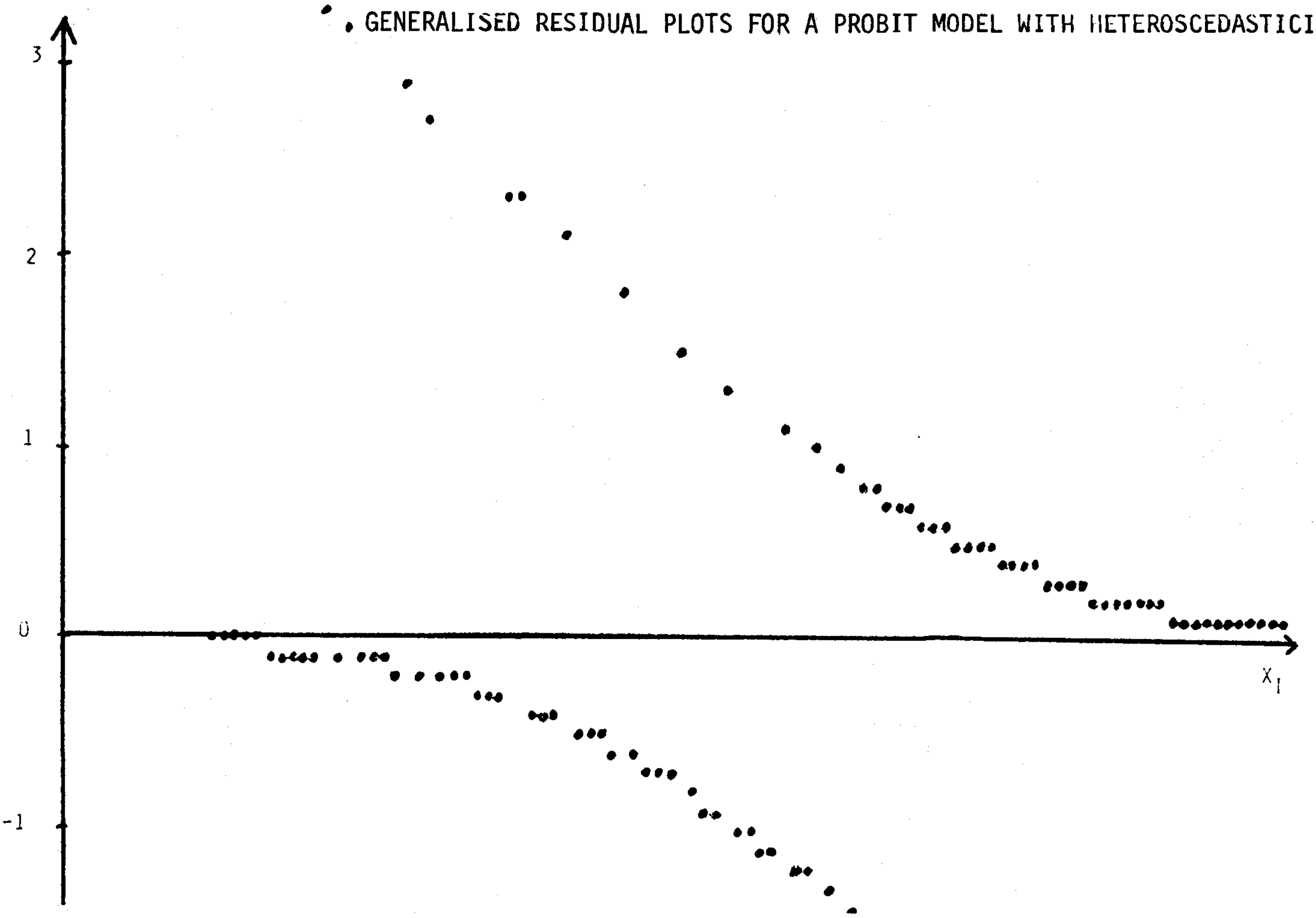


FIGURE 17

GENERALISED RESIDUAL PLOTS FOR A TOBIT MODEL WITH HETEROSCEDASTICITY

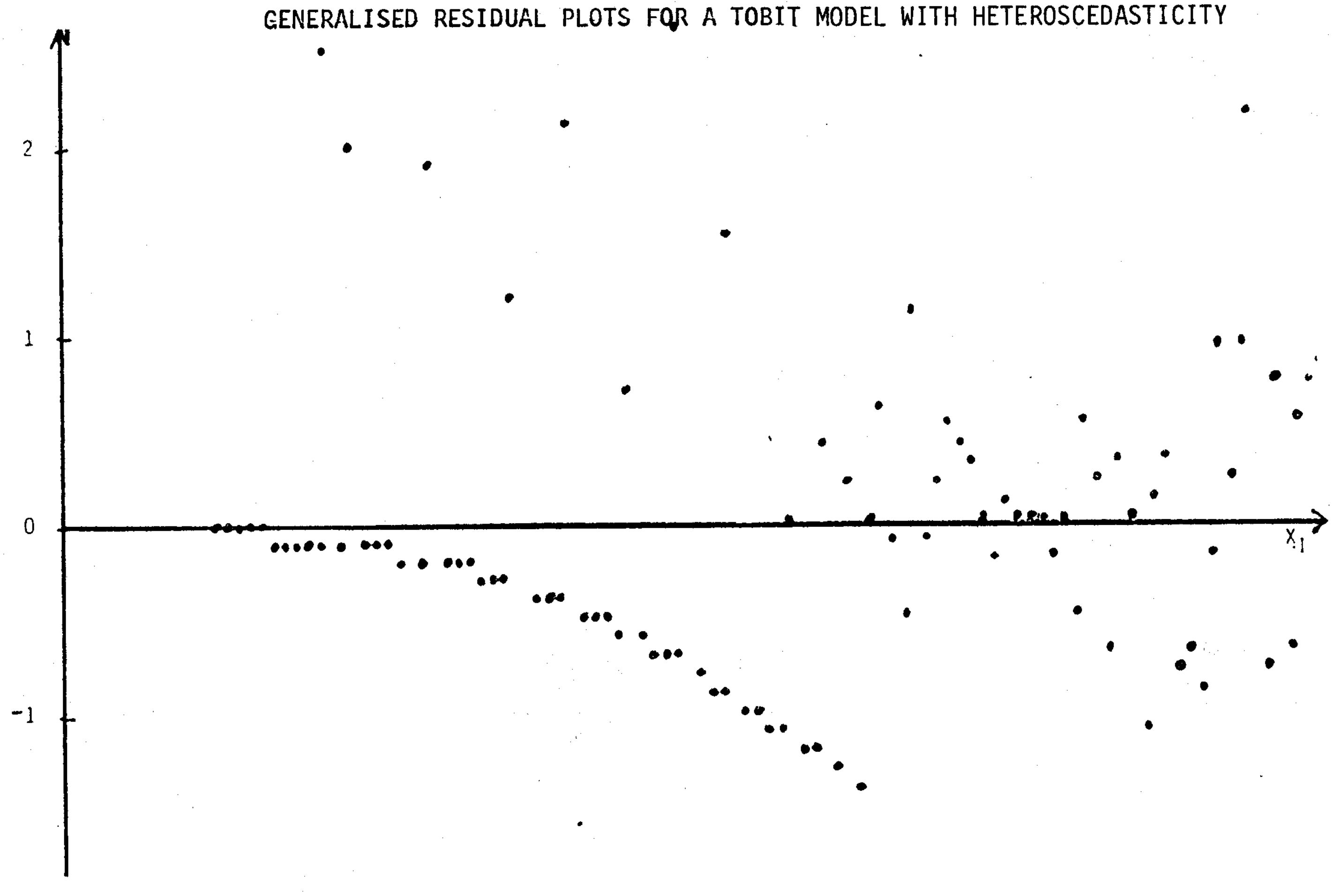


FIGURE 18

GENERALISED RESIDUAL PLOTS FOR A PROBIT MODEL WITH VARIABLES ALMOST UNCORRELATED WITH THE MAINTAINED EXOGENOUS VARIABLE

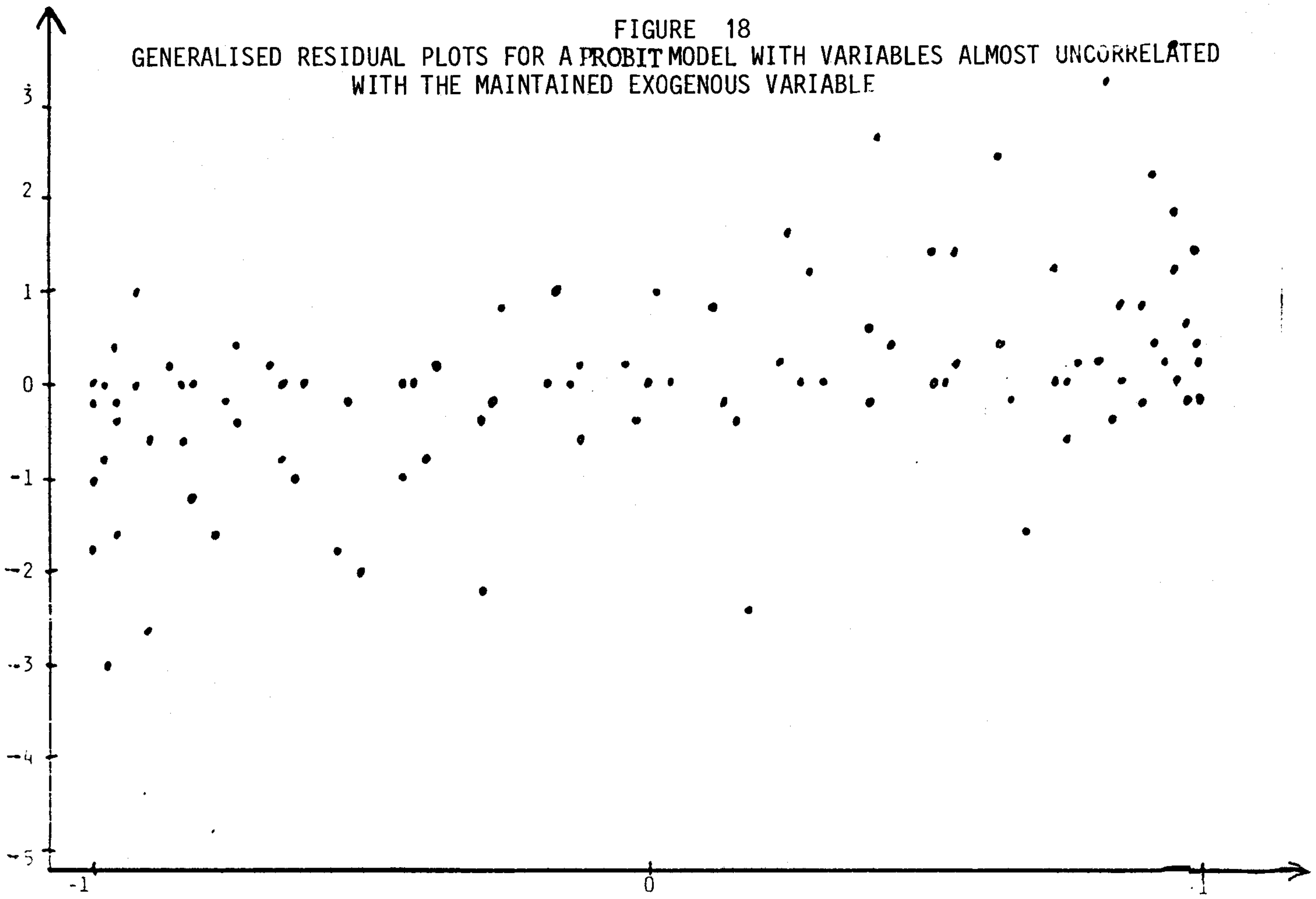


FIGURE 19

GENERALISED RESIDUAL PLOTS FOR A TOBIT MODEL WITH OMITTED VARIABLES ALMOST UNCORRELATED WITH MAINTAINED EXOGENOUS VARIABLE

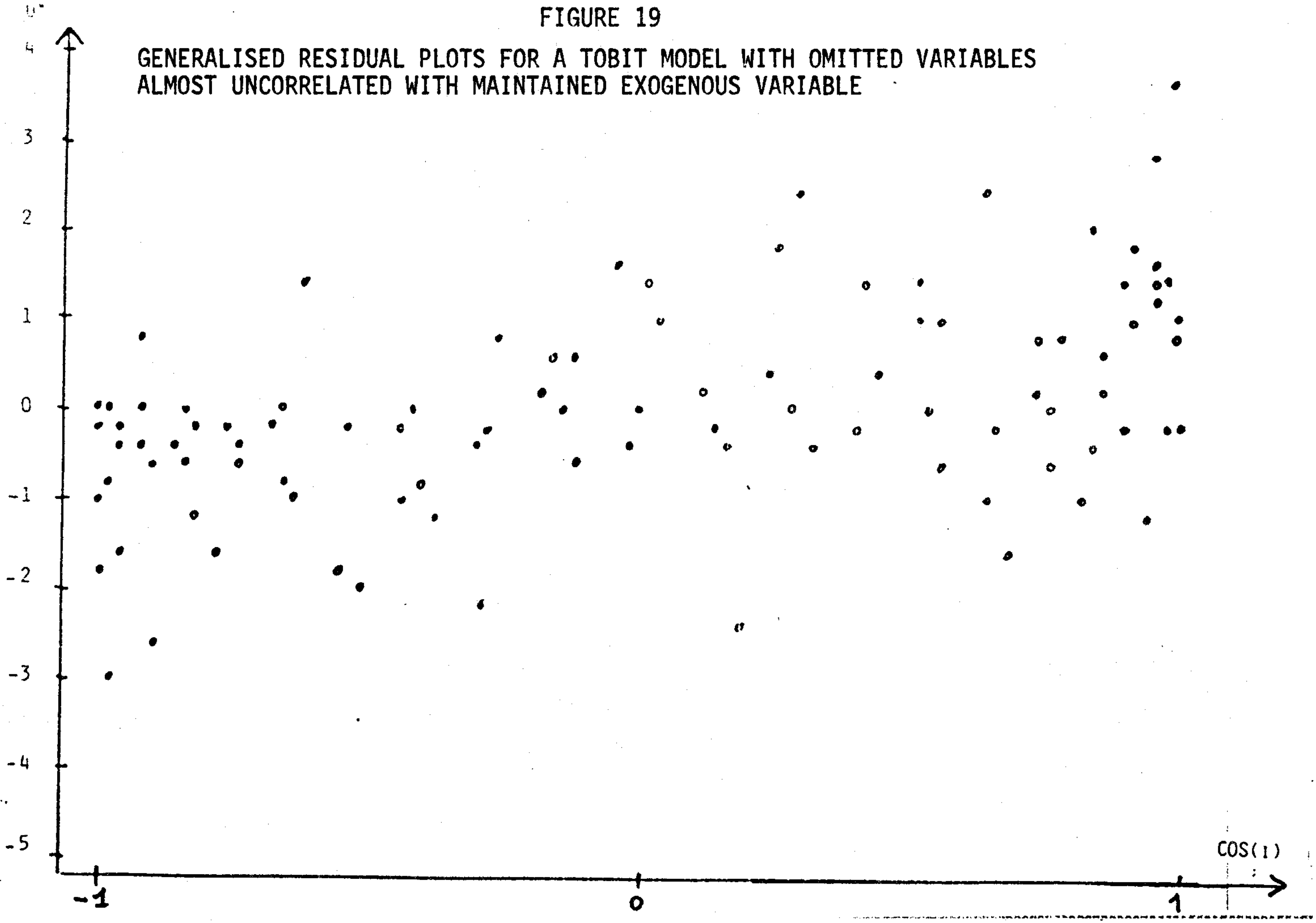


FIGURE 20

GENERALISED RESIDUAL PLOTS FOR A PROBIT MODEL WITH OMITTED VARIABLES HIGHLY CORRELATED WITH THE MAINTAINED EXOGENOUS VARIABLE

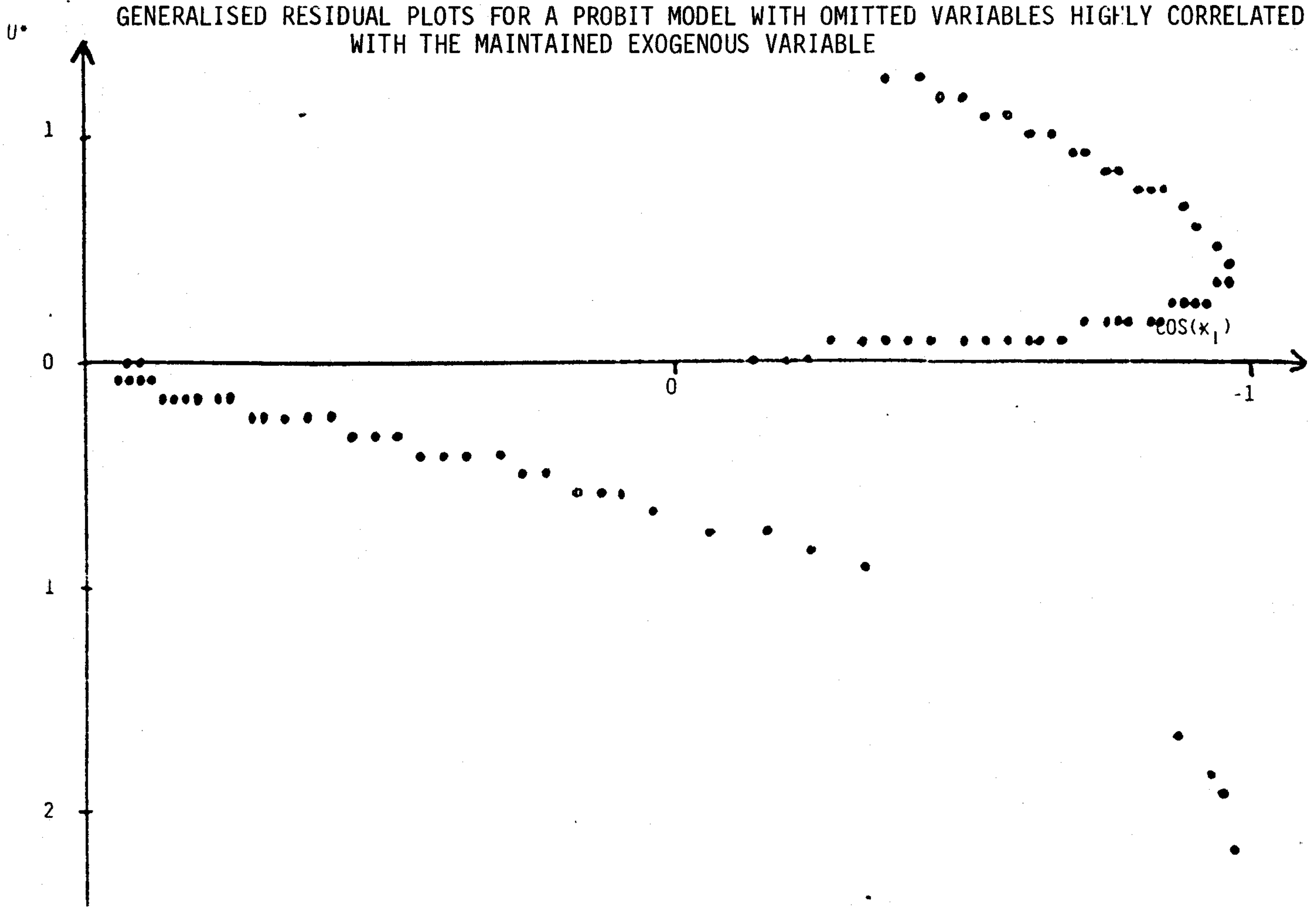
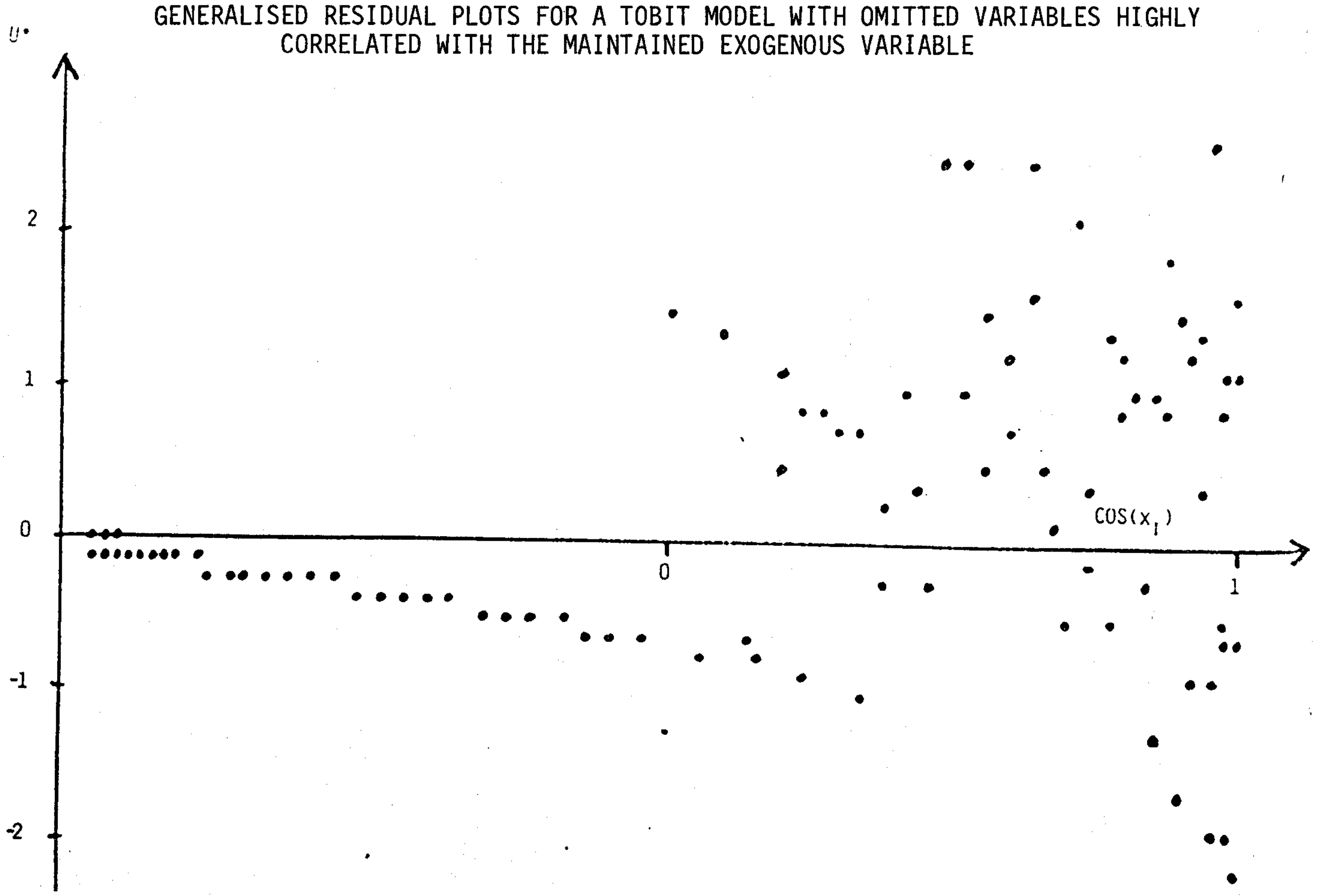


FIGURE 21

GENERALISED RESIDUAL PLOTS FOR A TOBIT MODEL WITH OMITTED VARIABLES HIGHLY CORRELATED WITH THE MAINTAINED EXOGENOUS VARIABLE



5 - TESTING PROCEDURES.

5.a - The case of a latent linear model.

Let us consider a linear model :

$$(5.1) \quad y_i^* = x_i b + u_i \qquad i = 1,\ldots,n$$

where the errors are i.i.d, normally distributed $N(o,\sigma^2)$. If y_i^* , x_i were observable, b would be estimated by OLS and some diagnostic tests would be performed. For instance, it would be possible to test for the significativity of some components of b or symmetrically to examine the relevance of some additional explanatory variables w_i . In the first case, the initial model (5.1) is the general hypothesis ; in the second case, model (5.1) is the null hypothesis and the general hypothesis has the form :

$$(5.2) \quad y_i^* = x_i b + w_i c + u_i$$

In this classical context the test procedures are based on Fisher statistics or equivalently on score statistics.

When the latent endogenous variable y^* is unobservable, the score statistics may be used after replacement of the unknown values y_i^* by the simulated ones z_{in} . The statistics thus obtained are called generalised score statistics. Depending on the hypothesis to be tested, the simulations are obtained under the null (case of omitted variables) or the general hypothesis (case of superfluous variables). The usual properties of the score statistics are no longer valid because of the replacement of y_i^* by z_{in} . In the following subsections, we are first interested in the determination of the right asymptotic covariance matrix of the generalised score statistic ; then by comparing this right asymptotic covariance matrix to the usual covariance matrix of the score statistic, we shall discuss the correct interpretation of the misspecified score test statistic, i.e the statistic in which the effect of the simulations has not been taken into account. In the case of a latent linear model, this misspecified statistic test simply corresponds to a F-test, applied to the regression of the z_{in} on the x_i .

5.b - Properties of the score statistic based on simulations under the null hypothesis.

The latent model is parameterised by a parameter θ , which can be partitioned in : $\theta = \begin{pmatrix}\alpha\\ \beta\end{pmatrix}$, where $\alpha \in R^{k_1}$, $\beta \in R^{k_2}$. We are interested in testing the nullity of the subparameter α ; the null hypothesis is given by : H_o : $(\alpha = o)$.
If the latent endogenous variable y^* were observable, the score statistic for testing Ho would be :

$$(5.3) \quad \xi_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \, \text{Log } l^*(y_i^*/x_i;o,\hat{\hat{\beta}}_{on})}{\partial\alpha}$$

where $\hat{\hat{\beta}}_{on}$ denotes the constrained maximum likelihood estimator of β . If the true value of the parameter belongs to the null and is : $\theta_o = \begin{pmatrix}o\\ \beta_o\end{pmatrix}$, ξ_n is asymptotically normal with zero mean and with a covariance matrix given by :

$$(5.4) \quad \boxed{V_{as}\,\hat{\xi}_n = I_{\alpha\alpha} - I_{\alpha\beta}\, I_{\beta\beta}^{-1}\, I_{\beta\alpha}}$$

where $I_{\alpha\alpha}$, $I_{\alpha\beta}$, $I_{\beta\beta}$ are the blocks of the information matrix I of the latent model evaluated at θ_o .

When the latent endogenous variable is unobservable,

i) the parameter θ can be estimated by the constrained maximum likelihood method applied to the observable model ;

ii) the estimator $\hat{\theta}_{on}$ thus obtained can be used as a basis for simulating the y_i^*. These simulations are denoted by z_{in} ;

iii) the simulated series is then used to compute a constrained maximum likelihood estimator $\tilde{\beta}_{on}$ of β. $\tilde{\beta}_{on}$ is a solution of :

$$\underset{\beta}{\text{Max}} \sum_{i=1}^{n} \text{Log } l^*(z_{in}/x_i;o,\beta) \quad ;$$

iv) the generalised score is obtained by replacing in the expression of $\hat{\xi}_n$ y_i^* by z_{in} and $\hat{\beta}_{on}$ by $\tilde{\beta}_{on}$.

The statistic is given by :

$$(5.5) \quad \boxed{\xi_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \text{ Log } l^*(z_{in}/x_i;o,\tilde{\beta}_{on})}{\partial\alpha}}$$

v) A misspecified test is based on the misspecified statistic $\xi_n'(\hat{V}_{as}\hat{\xi}_n)^{-1}\xi_n$, whose asymptotic properties have to be studied.

The asymptotic properties of ξ_n under the null are summarized in the following theorem.

<u>THEOREM 5.6</u> :

> Under a set of regularity conditions described in appendices 1, 2, 3, ξ_n is asymptotically normally distributed under the null, with zero mean and with an asymptotic covariance matrix given by :

$$V_{as}\,\xi_n = I_{\alpha\alpha} - I_{\alpha\beta}\, I^{-1}_{\beta\beta}\, I_{\beta\alpha}$$

$$- \; (I_{\alpha\beta}\, I^{-1}_{\beta\beta} - J_{\alpha\beta}\, J^{-1}_{\beta\beta})\; J_{\beta\beta}\; (I^{-1}_{\beta\beta}\, I_{\beta\alpha} - J^{-1}_{\beta\beta}\, J_{\beta\alpha})$$

where I and J denote the latent and observable information matrices evaluated at $\theta_o = \begin{pmatrix} o \\ \beta_o \end{pmatrix}$.

<u>Proof</u> : See appendix 3 . □

The correct asymptotic covariance matrix of the generalised statistic ξ_n has to be compared with the usual form of the covariance matrix :

(5.4) $\hat{V}_{as}\,\xi_n = I_{\alpha\alpha} - I_{\alpha\beta}\, I^{-1}_{\beta\beta}\, I_{\beta\alpha}$

It is directly deduced from the expression of $V_{as}\,\xi_n$ that $V_{as}\,\hat{\xi}_n$ is greater than $V_{as}\,\xi_n$ for the usual order on symmetric matrices : $\hat{V}_{as}\,\xi_n \gg V_{as}\,\hat{\xi}_n$.

This implies the following inequality between the right and misspecified "chi-square" statistics.

<u>Corollary 5.7</u> :

$$\xi'_n\,(\hat{V}_{as}\,\xi_n)^{-1}\,\xi_n \geq \xi'_n\,(\hat{V}_{as}\,\hat{\xi}_n)^{-1}\,\xi_n$$

where $\hat{V}_{as}\,\xi_n$ and $\hat{V}_{as}\,\hat{\xi}_n$ are consistent estimators of $V_{as}\,\xi_n$ and $\hat{V}_{as}\,\xi_n$.

If the asymptotic size is equal to 5 % , the correct asymptotic test consists :

- in rejecting the null if $\xi'_n\,(\hat{V}_{as}\,\xi_n)^{-1}\,\xi_n > \chi^2_{95\%}(K_1)$
- in accepting the null, otherwise.

The misspecified test procedure consists :

$$\begin{cases} \text{in rejecting the null if } \xi_n' (\hat{V}_{as}\, \hat{\xi}_n)^{-1}\, \xi_n > \chi^2_{95\%}(K_1) \\ \text{in accepting the null otherwise.} \end{cases}$$

From corollary 5.7, it is clear that whenever the null is rejected with the misspecified procedure, it is also rejected with the correct one. The misspecified procedure appears to favour the null, i.e the initial model. The asymptotic size of the misspecified test is less than 5 % ; such a test is sometimes called a conservative test.

In order to obtain an interpretation of the misspecified test, when it leads to the acceptation of the null, it is necessary to bound from below the right variance $V_{as}\, \xi_n$ by a function of the erroneous one $V_{as}\, \hat{\xi}_n$.

Property 5.8 :

We have :

$$V_{as}\, \xi_n \gg \frac{1}{\lambda} V_{as}\, \hat{\xi}_n$$

where λ is the maximum eigenvalue of $J_{\beta\beta}^{-1/2}\, I_{\beta\beta}\, J_{\beta\beta}^{-1/2}$, or $J_{\beta\beta}^{-1}\, I_{\beta\beta}$.

Proof : See appendix 4 . □

Since $I_{\beta\beta} - J_{\beta\beta} = J_{\beta\beta}^{1/2} \left[J_{\beta\beta}^{-1/2}\, I_{\beta\beta}\, J_{\beta\beta}^{-1/2} - Id \right] J_{\beta\beta}^{1/2}$ is non negative, all the eigenvalues of $J_{\beta\beta}^{-1/2}\, I_{\beta\beta}\, J_{\beta\beta}^{-1/2}$ are greater than one and in particular $\lambda > 1$.

If the value of the misspecified statistic leads to accept the null : $\xi_n' (\hat{V}_{as}\, \hat{\xi}_n)^{-1}\, \xi_n < \chi^2_{95\%}(K_1)$ and if moreover this value is smaller than $\frac{1}{\lambda} \chi^2_{95\%}(K_1)$, then the right statistic would also lead to accept the null.

In summary the misspecified test can be used as a test with three possible answers :

$$\begin{cases} \text{if } \xi_n' (\hat{V}_{as} \hat{\xi}_n)^{-1} \xi_n > \chi^2_{95\%}(K_1) & : \text{rejection} \\ \text{if } \xi_n' (\hat{V}_{as} \hat{\xi}_n)^{-1} \xi_n < \frac{1}{\hat{\lambda}} \chi^2_{95\%}(K_1) & : \text{acceptation} \\ \text{if } \frac{1}{\hat{\lambda}} \chi^2_{95\%}(K_1) < \xi_n' (\hat{V}_{as} \hat{\xi}_n)^{-1} \xi_n < \chi^2_{95\%}(K_1) & : \text{undetermination} \end{cases}$$

where $\hat{\lambda}$ is a consistent estimator of λ .

Remark 5.9 : In the limit case $J_{\beta\beta} = I_{\beta\beta}$, we also have $J_{\alpha\beta} = I_{\alpha\beta}$ using the nonnegativity of $I - J$; we see that $V_{as}\,\xi_n = V_{as}\,\hat{\xi}_n$ and that the two test procedures are asymptotically equivalent. This case appears when y is a sufficient statistic with respect to β .

Remark 5.10: It is easily seen from the proof given in appendix 4, that the bound $\frac{1}{\lambda}$ is the most accurate in particular when α is a unidimensional parameter.

Remark 5.11 : If the misspecified statistic does not immediately conclude in favour of the rejection, a good strategy would be to derive easily computable upper bound of λ (or $\hat{\lambda}$) such as the trace or the determinant of the matrix involved ; if it is not enough to get an answer one has to choose between the computation of $\hat{\lambda}$, with the risk of remaining in the undetermination area, and the computation of the correct statistic $\xi_n'(\hat{V}_{as}\,\xi_n)^{-1}\,\xi_n$, with the additional cost of computing terms such as $J_{\alpha\beta}$ which are not direct by-products of the maximisation of the likelihood function under H_0 .

5.c - Properties of the score statistic based on simulations under the general hypothesis.

The determination of the score statistic is similar to that of the previous subsection, except that at the first stage the parameter θ is estimated by the unconstrained maximum likelihood method applied to the observable model. This modification of the estimation procedure of θ implies a modification of the simulated values and of the constrained ML estimator of β . For sake of simplicity, we keep the same notations as before, simply adding a "bar" on z and $\tilde{\beta}$.

The score statistic is given by :

$$(5.12)\quad \bar{\xi}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \operatorname{Log} l^*(\bar{z}_{in}/x_i ; o, \bar{\tilde{\beta}}_{on})}{\partial \alpha}$$

THEOREM 5.13 :

Under a set of regularity conditions described in the appendices, $\bar{\xi}_n$ is asymptotically normally distributed under the null, with zero mean and with an asymptotic covariance matrix given by :

$$V_{as}\,\bar{\xi}_n = V_{as}\,\xi_n + (I^{\alpha\alpha})^{-1} J^{\alpha\alpha} (I^{\alpha\alpha})^{-1} - (J^{\alpha\alpha})^{-1}$$

where $(I^{\alpha\alpha})^{-1} = I_{\alpha\alpha} - I_{\alpha\beta}\,(I_{\beta\beta})^{-1}\, I_{\beta\alpha}$

$(J^{\alpha\alpha})^{-1} = J_{\alpha\alpha} - J_{\alpha\beta}\,(J_{\beta\beta})^{-1}\, J_{\beta\alpha}$

Proof : See appendix 5. □

In order to compare the right test statistic $\bar{\xi}'_n(\hat{V}_{as}\,\bar{\xi}_n)^{-1}\,\bar{\xi}_n$ with the misspecified one $\bar{\xi}'_n(\hat{V}_{as}\,\xi_n)^{-1}\,\bar{\xi}_n$, we have to compare the two asymptotic covariance matrices $V_{as}\,\bar{\xi}_n$ and $V_{as}\,\hat{\xi}_n$.

Property 5.14 :

The asymptotic covariance matrix $V_{as}\,\bar{\xi}_n$ is greater that $V_{as}\,\hat{\xi}_n$:

$$V_{as}\,\bar{\xi}_n >> V_{as}\,\xi_n \,.$$

Proof : It is possible by using an orthogonalisation procedure (see appendix 4, i)) to only consider the case $I_{\alpha\beta} = o$. Under this condition, we have :

$$V_{as}\bar{\xi}_n - V_{as}\hat{\xi}_n = -J_{\alpha\beta}J_{\beta\beta}^{-1}J_{\beta\alpha} + (I^{\alpha\alpha})^{-1}J^{\alpha\alpha}(I^{\alpha\alpha})^{-1} - (J^{\alpha\alpha})^{-1}$$

$$= I_{\alpha\alpha}J^{\alpha\alpha}I_{\alpha\alpha} - J_{\alpha\alpha}$$

Since $I >> J$, we have : $J^{-1} >> I^{-1}$ and in particular : $J^{\alpha\alpha} >> I_{\alpha\alpha}^{-1}$.

Then : $V_{as}\bar{\xi}_n - V_{as}\hat{\xi}_n = I_{\alpha\alpha}J^{\alpha\alpha}I_{\alpha\alpha} - J_{\alpha\alpha}$

$$>> I_{\alpha\alpha} - J_{\alpha\alpha} >> 0$$

□

Therefore : $\bar{\xi}_n'(\hat{V}_{as}\bar{\xi}_n)^{-1}\bar{\xi}_n < \bar{\xi}_n'(\hat{V}_{as}\hat{\xi}_n)^{-1}\bar{\xi}_n$ and whenever the null is accepted with the misspecified procedure, it is also accepted with the right one.

<u>Property 5.15</u> :

> We have the following inequality :
>
> $$V_{as}\bar{\xi}_n << V_{as}\hat{\xi}_n \left[1 + \mu - \frac{1}{\mu}\right]$$
>
> where μ is the maximum eigenvalue of $I^{1/2}J^{-1}I^{1/2}$, or $I J^{-1}$.

<u>Proof</u> : Considering without loss of generality the case $I_{\alpha\beta} = 0$, we have :

$$V_{as}\bar{\xi}_n = V_{as}\hat{\xi}_n + I_{\alpha\alpha}J^{\alpha\alpha}I_{\alpha\alpha} - J_{\alpha\alpha}$$

$$= V_{as}\hat{\xi}_n + I_{\alpha\alpha}^{1/2}\left[I_{\alpha\alpha}^{+1/2}J^{\alpha\alpha}I_{\alpha\alpha}^{+1/2} - I_{\alpha\alpha}^{-1/2}J_{\alpha\alpha}I_{\alpha\alpha}^{-1/2}\right]I_{\alpha\alpha}^{1/2}$$

$$<< V_{as}\hat{\xi}_n + I_{\alpha\alpha}^{1/2}(\mu\, Id - \frac{1}{\mu}Id)\, I_{\alpha\alpha}^{1/2}$$

$$= V_{as}\,\hat{\xi}_n + I_{\alpha\alpha}\left(\mu - \frac{1}{\mu}\right)$$

$$= V_{as}\,\hat{\xi}_n \left(1 + \mu - \frac{1}{\mu}\right)$$

□

Since $J^{-1} \gg I^{-1}$, $I^{1/2} J^{-1} I^{1/2} \gg Id$, the maximum eigenvalue μ is greater than one and $1 + \mu - \frac{1}{\mu}$ is also greater than 1. As in the previous subsection, the misspecified test may be considered as a test with three possible answers, but in this case this is the misspecified rejection region which is separate in a rejection and an undetermination region :

$$\begin{cases} \text{if } \bar{\xi}'_n (\hat{V}_{as}\,\hat{\xi}_n)^{-1} \bar{\xi}_n > \hat{\lambda}\chi^2_{95\%}(K_1) & : \text{rejection} \\ \text{if } \bar{\xi}'_n (\hat{V}_{as}\,\hat{\xi}_n)^{-1} \bar{\xi}_n < \chi^2_{95\%}(K_1) & : \text{acceptation} \\ \text{if } \chi^2_{95\%}(K_1) < \bar{\xi}'_n (\hat{V}_{as}\,\hat{\xi}_n)^{-1} \bar{\xi}_n < \hat{\lambda}\chi^2_{95\%}(K_1) & : \text{undetermination} \end{cases}$$

where $\hat{\lambda} = 1 + \hat{\mu} - \frac{1}{\hat{\mu}}$ and $\hat{\mu}$ is a consistent estimator of μ .

5.d - Asymptotic behaviour of the generalised Wald test and the generalised likelihood ratio test

Up to now we have only considered the generalised score test statistic, because the statistics $\frac{\xi_n}{\sqrt{n}}$ or $\frac{\xi_n}{\sqrt{n}}$ appear as sample means and, therefore, their asymptotic behaviour is easily tackled through the generalised central limit theorem 2.4. However it is natural to also consider the asymptotic behaviour of the generalised Wald test and of the generalised likelihood ratio test.

We consider the same genral framework as in 5.b and 5.c. The z_{in} are obtained from simulations based either on the constrained maximum likelihood of Θ or on the unconstrained maximum likelihood estimator of Θ . In the sequel we consider the constrained case but the other case would provide exactly the same result.

From the z_{in} we can compute unconstrained and constrained second stage estimators by maximising, respectively,

$$\sum_{i=n}^{n} \text{Log } l^*(z_{in}/x_i , \alpha, \beta) \text{ and } \sum_{i=1}^{n} \text{Log } l^*(z_{in}/x_i \ ; 0, \beta)$$

$\tilde{\theta}_n = \begin{pmatrix} \tilde{\alpha}_n \\ \tilde{\beta}_n \end{pmatrix}$ $\tilde{\theta}_{on} = \begin{pmatrix} 0 \\ \tilde{\beta}_{on} \end{pmatrix}$ will denote the unconstrained and the constrained estimator.

Using Taylor expansions and the law of large numbers previously shown we have, under H_0 :

$$\begin{pmatrix} \xi_n \\ 0 \end{pmatrix} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \text{ Log } l^*(z_{in}/x_i \ ; 0, \beta_{on}^2)}{\partial \ \theta} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \text{ Log } l^*(z_{in}/x_i \ ; 0, \beta_o)}{\partial \ \theta}$$

$$- I(\tilde{\theta}_o) \sqrt{n} \ (\tilde{\theta}_{no} - \theta_o) + o_p(1)$$

Substracting these expansions we get :

$$\begin{pmatrix} \xi_n \\ 0 \end{pmatrix} = I(\theta_o) \sqrt{n}(\tilde{\theta}_n - \tilde{\theta}_{no}) + o_p(1)$$

$$\sqrt{n} \ \tilde{\alpha}_n = I^{\alpha\alpha}(\theta_o) \ \xi_n + o_p(1)$$

It follows immediately that the generalised Wald test, based on $\tilde{\alpha}_n$, is asymptotically equivalent, under H_0 , to the generalised score test. As far as the generalised likelihood ratio test is concerned, it can easily be shown, using the standard expansions that :

$$2[\sum_{i=1}^{n} \text{Log } l^*(z_{in}/x_i \ ; \tilde{\alpha}_n, \tilde{\beta}_n) - \sum_{i=1}^{n} \text{Log } l^*(z_{in}/x_i \ ; 0, \tilde{\beta}_{on}] = \xi_n' \ I^{\alpha\alpha}(\theta_o) \ \xi_n + o_p(1)$$

Since $I^{\alpha\alpha}(\theta_o)$ is not the inverse of the asymptotic covariance matrix of ξ_n , this generalised likelihood ratio is not asymptotically distributed as a chi-square but as a mixture of chi-squares [see Foutz-Srivastava (1977)]. In particular, this statistic is not asymptotically equivalent to the generalised score and wald statistic ; the situation is similar to the one found in the pseudo-likelihood theory (see TROGNON (1983), GOURIEROUX-MONFORT-TROGNON (1984b)).

5.e - Asymptotic equivalence of the misspecified score, Wald and likelihood ratio test statistics, in the case of a linear latent model

In the previous subsections, we have discussed the correct asymptotic properties of generalised statistics and the correct interpretation of the misspecified score test statistics. In practice the usual estimation packages also provide the values of some "misspecified" Wald type statistics (e.g. student or Fisher statistics) and of the maximum of log-likelihood function. Considering the case of a latent linear model, we are going to prove that these three kinds of misspecified statistics are asymptotically equivalent under the null and in particular that the level correction procedure described in 5.b, 5.c is valid for all these tests procedures.

Let us consider the following latent linear model :

$$(5.16)\quad y_i^* = x_i\beta_o + w_i\alpha_o + u_i \qquad i = 1,\dots,n\ ,\ u_i \text{ i.i.d, } u_i \approx N(o,\sigma_o^2)$$

and the null hypothesis given by :

$$Ho = \left\{\alpha = o\right\} \quad .$$

Since y_i^* is unobservable, the observations y_i^* of the latent endogenous variable are replaced by the simulated values z_{in} and the initial model (5.16) by the misspecified one :

$$(5.17)\quad z_{in} = x_i\beta + w_i\alpha + u_i \qquad i = 1,\dots,n\ ,\ u_i \text{ i.i.d, } u_i \approx N(o,\sigma^2)$$

The misspecified test statistics of the null hypothesis are given by the usual formulae. Let us denote by M_X (resp $M_{X,W}$) the orthogonal projector on the space orthogonal to that generated by the columns of X (resp. of X , W) and by $\tilde{\alpha}$ the unconstrained o.l.s. estimator of α based on the misspecified model :

$$\tilde{\alpha} = [W'M_XW]^{-1}\ W'M_X Z$$

The misspecified test statistics are :

$$\text{score test statistic : } \quad n\,\frac{\tilde{\alpha}'\ W'M_XW\ \tilde{\alpha}}{Z'M_XZ}$$

Wald statistic :
$$n \frac{\tilde{\alpha}' W' M_X W \tilde{\alpha}}{Z' M_{X,W} Z}$$

Likelihood ratio statistic :
$$n \operatorname{Log} \frac{Z' M_X Z}{Z' M_{X,W} Z}$$

$$= n \operatorname{Log} \left[1 + \frac{\tilde{\alpha}' W' M_X W \tilde{\alpha}}{Z' M_{X,W} Z} \right]$$

Then the asymptotic equivalence of these statistics under the null is simply a consequence of the law of large numbers (2.6), which implies that, under the null :

$$\operatorname{plim} \frac{1}{n} Z' M_X Z = \operatorname{plim} \frac{1}{n} Y^{*'} M_X Y^* = \sigma_o^2$$

$$\operatorname{plim} \frac{1}{n} Z' M_{X,W} Z = \operatorname{plim} \frac{1}{n} Y^{*'} M_{X,W} Y^* = \sigma_o^2 .$$

and of the asymptotic properties of $\tilde{\alpha}$.

For instance, let us consider the probit model of section 1. From the values of the information matrices I and J given in 3-b , we obtain an estimate $\hat{\mu}$ of μ equal to 5.34 .

If we want to test at level 5 % the null hypothesis H_0 : $(\beta = 1)$ using the misspecified student statistic, we have the following three answers Wald test :

If the t-statistic has a modulus smaller than 1.96, we accept the null hypothesis.

If the t-statistic has a modulus greater than :

$1.96 \sqrt{1 + \hat{\mu} - \frac{1}{\hat{\mu}}} = 4.86$, we reject the null hypothesis.

Otherwise the test is inconclusive.

The observed values of the t-statistic is -2.044 and the last answer is the right one.

CONCLUDING REMARKS

The results proved in this paper, though they are not trivial from a technical point of view, are very easy to use in practice. In particular, if the latent model is linear, the basic tools are simply the usual residuals provided by standard regression packages. This implies that various graphical checks or genuine statistical tests can easily be implemented for various models such as probit, Tobit, disequilibrium models... These new possibilities seem to be important since, in these kinds of models, it is well-known that specifications errors may have much more serious consequences than in the usual linear model.

APPENDIX 1

PROOF OF THE GENERALISED CENTRAL LIMIT THEOREM.

i) The theorem

Let (X_i, Y_i^*) , $i \in N$ be i.i.d random variables whose values are in $R^{d_o+d_1}$. The common distribution of these pairs belongs to a parametrised family : $\langle P_\theta , \theta \in \Theta \subset R^k \rangle$ and is associated with the value θ_o of the parameter.

For a given function g from R^{d_1} into R^{d_2} , we define the transformed variables: $Y_i = g(Y_i^*)$, $i \in N$. If the observed variables are (X_i , Y_i) $i = 1...n$, an estimator $\hat{\theta}_n$ of θ is a measurable function of (X_i , Y_i) $i = 1...n$, with values in R^K . From this estimator, it is possible to generate approximations of the latent variables Y_i^* by drawing independently from the conditional distribution of Y_i^* given Y_i and X_i , associated with the value $\hat{\theta}_n$ of the parameter ; these simulated series are denoted by Z_{in} , $i = 1...n$, $n \in N$.

We are now interested in the asymptotic properties of an empirical mean :

$$\frac{1}{n} \sum_{i=1}^{n} h(Z_{in}, Y_i, X_i) \text{ , where h is a H-dimensional function.}$$

Theorem 2.4 : Let us assume that :

$$\sqrt{n}(\hat{\theta}_n - \theta_o) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} a(Y_i, X_i) + \varepsilon_{1n}$$

with a function a satisfying :

$$E_{\theta_o} ||a(Y,X)||^2 < \infty \ , \ E_{\theta_o} \ a(Y,X) = 0$$

and a random term ε_{1n} tending in probability to zero :

$$\varepsilon_{1n} \xrightarrow[n\to\infty]{P_{\theta_o}} 0 \ .$$

Then, under the regularity conditions given below, the random variable :

$$\xi_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[h(Z_{in},Y_i,X_i) - E_{\theta_o} h(Y^*,Y,X) \right]$$

tends in distribution to a normal distribution with zero mean and a covariance matrix given by :

$$V_{as}\xi = V_{\theta_o} [h(Y^*,Y,X)] - V_{\theta_o}\left[E_{\theta_o} (h(Y^*,Y,X)/Y,X)\right]$$

$$+ V_{\theta_o}\left\{E_{\theta_o} \left[h(Y^*,Y,X)/Y,X\right] + E_{\theta_o}\left\{\left[\frac{\partial}{\partial\theta'} E_{\theta}\left[h(Y^*,Y,X)/Y,X\right]_{\theta=\theta_o} a\ (Y,X)\right]\right.\right.$$

ii) <u>Regularity conditions</u>

H1 : θ_o belongs to the interior of Θ .

H2 : The marginal distribution of X_i is independent from θ , $\theta \in \Theta$.

H3 : The conditional distribution $P_\theta^{Y^*/X=x}$ has a density with respect to $P_{\theta_o}^{Y^*/X=x}$ for any x and this density function is strictly positive. It is denoted by $l(y^*/x;\theta)$.

H4 : This density function is continuous at θ_o .

H5 : $E_{\theta_o} |t'h(Y^*,Y,X)|^4 < +\infty \qquad \forall\ t \in R^H$

H6 : $\forall\, t, \forall\, k, \exists\, n_o(k) : \forall\, n > n_o(k)\ \forall\, \alpha = 1,2,3,4$

$$E_{\theta_o}\left(|t'\left[h(Y^*,Y,X) - E_{\theta_o}(Y^*,Y,X)\right]|^{\alpha} \sup_{\theta : ||\theta-\theta_o|| < \frac{k}{\sqrt{n}}} |l(Y^*/Y,X;\theta)-1| \right)$$

where $l(y^*/y,x;\theta)$ denotes the density function of $P_{\theta}^{Y^*/Y,X}$ with respect to $P_{\theta_o}^{Y^*/Y,X}$.

H7 : $E_{\theta}\,[t'h(Y^*,Y,X)/Y,X]$ is differentiable with respect to θ and the first derivative is continuous at θ_o, for any $t \in R^H$.

H8 : $E_{\theta_o} \left|\left| \frac{\partial}{\partial\theta} E_{\theta_o}\left[t'h(Y^*,Y,X)/Y,X\right] \right|\right| < +\infty$, $\forall\, t \in R$.

H9 : For any t, there exists $c_o > 0$ such that :

$$E_{\theta_o} \sup_{\theta : ||\theta-\theta_o|| < c_o} \left|\left| \frac{\partial}{\partial\theta} E_{\theta}\left[t'h(Y^*,Y,X)/Y,X\right] - \frac{\partial}{\partial\theta} E_{\theta_o}\left[t'h(Y^*,Y,X)/Y,X\right] \right|\right| < +\infty$$

iii) A lemma.

In the proof of the theorem we use several times the following lemma :

Lemma : Let $(\alpha_i, \Psi_i, \Psi_i^*)$ $i = 1 \ldots n$ be random variables such that :

(i) α_i (resp Ψ_i, Ψ_i^*) takes its values in R^+ (resp in C) ;

(ii) the random variables α_i are independent and integrable ;

(iii) $|\Psi_i| < 1$ and $|\Psi_i - \Psi_i^*| < \alpha_i$ $\forall\, i = 1 \ldots n$; then

$$|E(\prod_{i=1}^{n} \Psi_i^*) - E(\prod_{i=1}^{n} \Psi_i)| \leq \exp(\sum_{i=1}^{n} E\,\alpha_i) - 1$$

<u>Proof of the lemma</u> : Let us first remark that, since $|\Psi_i| < 1$, the product $\prod_{i=1}^{n} \Psi_i$ is integrable. Denoting $\beta_i = \Psi_i^* - \Psi_i$, we have :

$$\prod_{i=1}^{n} \Psi_i^* - \prod_{i=1}^{n} \Psi_i = \prod_{i=1}^{n} (\Psi_i + \beta_i) - \prod_{i=1}^{n} \Psi_i = \sum_{p=1}^{n} \sum_{A\in \mathcal{A}_p} \left(\prod_{i\in A} \beta_i\right) \left(\prod_{i\notin A} \Psi_i\right)$$

where $\mathcal{A}_p$ is the family of the subsets of $\langle 1...n\rangle$ with p elements

Taking the absolute values of each member of the equation, we obtain the following inequality :

$$|\prod_{i=1}^{n} \Psi_i^* - \prod_{i=1}^{n} \Psi_i| \leqslant \sum_{p=1}^{n} \sum_{A\in \mathcal{A}_p} \prod_{i\in A} |\beta_i| \prod_{i\notin A} |\Psi_i|$$

$$\leqslant \sum_{p=1}^{n} \sum_{A\in \mathcal{A}_p} \prod_{i\in A} |\beta_i| \qquad \text{(from (iii))}$$

$$\leqslant \sum_{p=1}^{n} \sum_{A\in \mathcal{A}_p} \prod_{i\in A} \alpha_i \qquad \text{(from (iii))}$$

Since the random variables α_i are independent, we have :

$$E|\prod_{i=1}^{n} \Psi_i^* - \prod_{i=1}^{n} \Psi_i| \leqslant \sum_{p=1}^{n} \sum_{A\in \mathcal{A}_p} \prod_{i\in A} E\alpha_i$$

Therefore : $E|\prod_{i=1}^{n} \Psi_i^*| \leqslant E|\prod_{i=1}^{n} \Psi_i| + \sum_{p=1}^{n} \sum_{A\in \mathcal{A}_p} \prod_{i\in A} E\alpha_i < \infty$

which means that $\prod_{i=1}^{n} \Psi_i^*$ is integrable and we have :

$$|E\prod_{i=1}^{n} \Psi_i^* - E\prod_{i=1}^{n} \Psi_i| \leqslant E|\prod_{i=1}^{n} \Psi_i^* - \prod_{i=1}^{n} \Psi_i|$$

$$\leq \sum_{p=1}^{n} \sum_{A \in \mathcal{A}_p} \prod_{i \in A} (E\alpha_i)$$

$$= \prod_{i=1}^{n} (1 + E\alpha_i) - 1$$

$$\leq \exp \sum_{i=1}^{n} E\alpha_i - 1$$

Q.E.D.

iv) Proof of the theorem :

The asymptotic distribution of ξ_n is deduces from a study of its characteristic function.

a) Expression of the characteristic function.

We have :

$$E_{\theta_o} \exp(jt'\xi_n) = E_{\theta_o} \left\{ \exp \frac{jt'}{\sqrt{n}} \sum_{i=1}^{n} [h(Z_{in}, Y_i, X_i) - E_{\theta_o} h(Y^*, Y, X)] \right\}$$

$$= E_{\theta_o} E_{\theta_o} \left[\prod_{i=1}^{n} \exp \frac{jt'}{\sqrt{n}} [h(Z_{in}, Y_i, X_i) - E_{\theta_o} h(Y^*, Y, X)] / (X_1, Y_1) \ldots (X_n, Y_n) \right]$$

Since the simulated observations are drawn independently from the conditional distribution of Y_i^* given Y_i, X_i associated with the value $\hat{\theta}_n$ of the parameter, we deduce :

$$E_{\theta_o} [\exp jt' \xi_n] = E_{\theta_o} [\prod_{i=1}^{n} \varphi_{in}]$$

with $\varphi_{in} = E_{\hat{\theta}_n} \left[\exp \frac{jt'}{\sqrt{n}} [h(Y_i^*,Y_i,X_i) - E_{\theta_o} h(Y^*,Y,X)] / X_i,Y_i\right]$

We are now going to asymptotically expand the expression of φ_{in}. For this purpose, we define for each pair of integers (n,k), the subset

$$\Omega_{n,k} = \left[||\hat{\theta}_n - \theta_o|| < \frac{k}{\sqrt{n}} \right] \cap \left[||\varepsilon_{1n}|| < \frac{1}{k} \right] .$$

b) <u>Expansion of φ_{in} on $\Omega_{n,k}$</u>.

b.i) If x is a real number, we have :

$$|\exp j x - 1 - j x + \frac{x^2}{2} | \leq \frac{|x|^3}{3!} + \frac{x^4}{4!}$$

Therefore, denoting $h_i = h(Y_i^*,Y_i,X_i)$, we get :

$$\exp j \frac{t'}{\sqrt{n}} (h_i - E_{\theta_o} h) = 1 + \frac{j}{\sqrt{n}} t'(h_i - E_{\theta_o} h) - \frac{1}{2n} \left[t'(h_i - E_{\theta_o} h)\right]^2$$

$$+ o_{1n}(i)$$

$$\text{with } |o_{1n}(i)| \leq \frac{1}{4!} \left[\frac{t'(h_i - E_{\theta_o} h)}{\sqrt{n}} \right]^4 + \frac{1}{3!} \left| \frac{t'(h_i - E_{\theta_o} h)}{\sqrt{n}} \right|^3$$

b.ii) From assumptions H5 and H6, we deduce that $E_{\theta} |t'(h(Y^*,Y,X) - E_{\theta_o} h)|^{\alpha} < \infty$ for $\alpha = 1,2,3,4$ and for θ satisfying : $||\theta - \theta_o|| < \frac{k}{\sqrt{n}}$.

Therefore the conditional expectations

$$E_{\theta}\left[\,|t'(h(Y^*,Y_i,X_i) - E_{\theta_o} h)|^{\alpha} \Big/ Y_i,X_i\,\right]$$

exist under the same conditions. This implies that on the subset $\Omega_{n,k}$, the expectations $E_{\hat{\theta}_n}\left[\,|t'(h(Y^*,Y_i,X_i) - E_{\theta_o} h)|^{\alpha} \Big/ Y_i,X_i\,\right]$

have a sense.

Then it is possible to express the general term φ_{in} as :

$$\varphi_{in} = E_{\hat{\theta}_n}\left[\exp j \frac{t'}{\sqrt{n}} (h_i - E_{\theta_o} h) \Big/ Y_i,X_i\right]$$

$$= 1 + \frac{jt'}{\sqrt{n}} E_{\hat{\theta}_n}\left[t'(h_i - E_{\theta_o} h)| \Big/ Y_i,X_i\right]$$

$$- \frac{1}{2n} E_{\hat{\theta}_n}\left[\left[t'(h_i - E_{\theta_o} h)\right]^2 \Big/ Y_i,X_i\right] + E_{\hat{\theta}_n} o_{1n}(i)$$

$$= 1 + \frac{jt'}{\sqrt{n}} E_{\hat{\theta}_n}\left[t'(h_i - E_{\theta_o} h)| \Big/ Y_i,X_i\right]$$

$$- \frac{1}{2n} E_{\hat{\theta}_n}\left[\left[t'(h_i - E_{\theta_o} h)\right]^2 \Big/ Y_i,X_i\right] + o_{2n}(i)$$

The residual term $o_{2n}(i)$ depends on all values $X_1,\dots,X_n,Y_1,\dots,Y_n$ through the estimator $\hat{\theta}_n$ of θ_o. However, it is smaller on $\Omega_{n,k}$ than a random variable depending only on X_i,Y_i. More precisely, if we denote by $\eta_{k,n}(Y^*,Y_i,X_i) = \eta_{k,n}(i)$ the quantity $\sup_{\theta:||\theta-\theta_o||<\frac{k}{\sqrt{n}}} |l(Y^*/Y_i,X_i;\theta)-1|$

we obtain on the subset $\Omega_{n,k}$:

$$| o_{2n}(i)| \leq \bar{o}_{2n}(Y_i,X_i)$$

with : $$\bar{o}_{2n}(Y_i,X_i) = \frac{1}{2n} E_{\theta_o}\left[\left[t'(h_i - E_{\theta_o} h)\right]^2 \eta_{k,n}(i)/Y_i X_i)\right]$$

$$+ \frac{1}{4!n^2} E_{\theta_o}\left[\left[t'(h_i - E_{\theta_o} h)\right]^4 / Y_i,X_i\right]$$

$$+ \frac{1}{3!n^{3/2}} E_{\theta_o}\left[\left[t'(h_i - E_{\theta_o} h)\right]^3 / Y_i,X_i\right]$$

b.iii) Since $\hat{\theta}_n$ tends to θ_o, it is natural to expand the second term of the expression of φ_{in} :

$$E_{\hat{\theta}_n}\left[t'(h_i - E_{\theta_o} h)| \Big/ Y_i,X_i\right] = E_{\theta_o}\left[t'(h_i - E_{\theta_o} h)| \Big/ Y_i,X_i\right]$$

$$+ \left[\frac{\partial}{\partial\theta'} E_{\theta_o}(t'h_i/Y_i,X_i)\right] (\hat{\theta}_n - \theta_o) + o_{3n}(i)$$

The residual term $o_{3n}(i)$ is on $\Omega_{n,k}$ smaller than :

$$|o_{3n}(i)| \leq ||\hat{\theta}_n - \theta_o|| \sup_{\theta\in[\theta_o,\hat{\theta}_n]} ||\frac{\partial}{\partial\theta} E_{\theta}(t'h_i/Y_i,X_i) - \frac{\partial}{\partial\theta} E_{\theta_o}(t'h_i/Y_i,X_i)||$$

$$\leq \frac{k}{\sqrt{n}} \sup_{\theta:||\theta-\theta_o||\frac{k}{\sqrt{n}}} ||\frac{\partial}{\partial\theta} E_{\theta}(t'h_i/Y_i,X_i) - \frac{\partial}{\partial\theta} E_{\theta_o}(t'h_i/Y_i,X_i)||$$

$$= \bar{o}_{3n}(Y_i,X_i) \quad \text{(say)}$$

b.iv) Using the asymptotic expansion of $\hat{\theta}_n$ and the fact that $||\varepsilon_{in}|| < \frac{1}{k}$ on $\Omega_{n,k}$, we next obtain :

$$E_{\hat{\theta}_n}\left(t'(h_i - E_{\theta_o} h)/Y_i, X_i\right) = E_{\theta_o}\left(t'(h_i - E_{\theta_o} h)/Y_i, X_i\right)$$

$$+ \frac{\partial}{\partial\theta'} E_{\theta_o}\left(t' h_i/Y_i, X_i\right) \cdot \frac{1}{n}\sum_{i=1}^{n} a(Y_i, X_i) + o_{3n}(i) + o_{4n}(i)$$

with $|o_{4n}(i)| = \left| \frac{\partial}{\partial\theta'} E_{\theta_o}\left(t' h_i/Y_i X_i\right) \frac{\varepsilon_{in}}{\sqrt{n}} \right|$

$$< \frac{1}{k\sqrt{n}} \left|\left| \frac{\partial}{\partial\theta'} E_{\theta_o}\left(t' h_i/Y_i X_i\right)\right|\right| = \bar{o}_{4n}(Y_i, X_i)$$

b.v) Replacing in the expression of φ_{in}, we get :

$$\varphi_{in} = \varphi^*_{in} + o_{2n}(i) + \frac{j}{\sqrt{n}} o_{3n}(i) + \frac{j}{\sqrt{n}} o_{4n}(i)$$

where : $\varphi^*_{in} = 1 + \frac{j}{\sqrt{n}} E_{\theta_o}\left(t'(h_i - E_{\theta_o} h)/Y_i X_i\right)$

$$+ \frac{j}{n}\frac{\partial}{\partial\theta'} E_{\theta_o}\left(t'h_i/Y_i X_i\right) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} a(Y_i, X_i)$$

$$- \frac{1}{2n} E_{\theta_o}\left\{\left(t'(h_i - E_{\theta_o} h)\right)^2/Y_i, X_i\right\}$$

Therefore, for n sufficiently large:

$$|\mathbb{1}_{\Omega_{n,k}} \varphi_{in} - \mathbb{1}_{\Omega_{n,k}} \varphi^*_{in}| < \bar{o}_{n,k}(Y_i, X_i) = \bar{o}_{2n}(Y_i, X_i) + \frac{1}{\sqrt{n}} \bar{o}_{3n}(Y_i, X_i) + \frac{1}{\sqrt{n}} \bar{o}_{4n}(Y_i, X_i)$$

c) <u>Comparison of</u> $1\!\!1_{\Omega_{n,k}} \prod_{i=1}^{n} \varphi_{in}$ <u>and</u> $1\!\!1_{\Omega_{n,k}} \prod_{i=1}^{n} \varphi_{in}^{*}$.

c.i) Noting that $\bar{o}_{n,k}(Y_i,X_i)$ is integrable, a direct application of the lemma given at the beginning, with $\Psi_{in} = 1\!\!1_{\Omega_{n,k}} \varphi_{in}$, $\Psi_{in}^{*} = 1\!\!1_{\Omega_{n,k}} \varphi_{in}^{*}$, $\alpha_i = \bar{o}_{n,k}(Y_i,X_i)$ gives :

$$E_{\theta_o} \left| 1\!\!1_{\Omega_{n,k}} \prod_{i=1}^{n} \varphi_{in} - 1\!\!1_{\Omega_{n,k}} \prod_{i=1}^{n} \varphi_{in}^{*} \right| \leq \exp\left(\sum_{i=1}^{n} E_{\theta_o} \bar{o}_{n,k}(Y_i,X_i) \right) - 1$$

$$= \exp \left[n \, E_{\theta_o} \bar{o}_{n,k}(Y_i,X_i) \right] - 1$$

c.ii) If we examine the decomposition of $n \; \bar{o}_{2n}(Y_i,X_i)$, it is easily seen that $\lim_{n\to\infty} n \; \bar{o}_{2n}(Y_i,X_i) = 0$, since $\lim_{n\to\infty} \eta_{k,n}(i) = 0$ from H4 . Using the dominated convergence theorem, we deduce that :

$$\lim_{n\to\infty} E_{\theta_o} \bar{o}_{2n}(Y_i,X_i) = 0$$

c.iii) Similarly, we have :

$$E_{\theta_o} \left(n \, \frac{\bar{o}_{3n}(Y_i,X_i)}{\sqrt{n}} \right) = k \; E_{\theta_o} \sup_{\theta : ||\theta-\theta_o|| < \frac{k}{\sqrt{n}}} \left|\left| \frac{\partial}{\partial\theta} E_{\theta} \left(t'h_i / Y_i, X_i \right) - \frac{\partial}{\partial\theta} E_{\theta_o} \left(t'h_i / Y_i, X_i \right) \right|\right|$$

Using H7 and H9 , we deduce from the dominated convergence theorem that :

$$\lim_{n\to\infty} E_{\theta_o}\left(n \frac{\bar{o}_{3n}(Y_i,X_i)}{\sqrt{n}} \right) = 0 \quad .$$

c.iv) Finally :

$$E_{\theta_o}\left(n \frac{\bar{o}_{4n}(Y_i,X_i)}{\sqrt{n}} \right) = \frac{1}{k} E_{\theta_o} \left(\left\| \frac{\partial}{\partial\theta} E_{\theta_o} \left(t' h_i / Y_i, X_i \right) \right\| \right)$$

c.v) Therefore :

$$\limsup_{n\to\infty} E_{\theta_o} \left| 1\!\!1_{\Omega_{n,k}} \prod_{i=1}^{n} \varphi_{in} - 1\!\!1_{\Omega_{n,k}} \prod_{i=1}^{n} \varphi^*_{in} \right|$$

$$\leq \exp\left[\frac{1}{k} E_{\theta_o} \left\| \frac{\partial}{\partial\theta} E_{\theta_o} (t' h_i / Y_i, X_i) \right\| \right] - 1$$

Finally, using H8 , we get :

$$\boxed{\limsup_{k\to\infty} \limsup_{n\to\infty} E_{\theta_o} \left| 1\!\!1_{\Omega_{n,k}} \prod_{i=1}^{n} \varphi_{in} - 1\!\!1_{\Omega_{n,k}} \prod_{i=1}^{n} \varphi^*_{in} \right| = 0}$$

d) <u>Comparison of</u> $1\!\!1_{\Omega_{n,k}} \prod_{i=1}^{n} \varphi^*_{in}$ <u>and</u> $1\!\!1_{\Omega_{n,k}} \prod_{i=1}^{n} \varphi^{**}_{in}$.

d.i) Let us define the random variable

$$\varphi^{**}_{in} = \exp(j A_{in} + B_{in})$$

where : $A_{in} = \frac{1}{\sqrt{n}} \, E_{\theta_o} \left(t'(h_i - E_{\theta_o} h)/Y_i, X_i \right)$

$$+ \frac{1}{n} \frac{\partial}{\partial \theta'} E_{\theta_o} \left(t'h_i/Y_i, X_i \right) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} a(Y_i, X_i)$$

$$B_{in} = - \frac{1}{2n} E_{\theta_o} \left\{ \left(t'(h_i - E_{\theta_o} h) \right)^2 / Y_i, X_i \right\}$$

$$+ \frac{1}{2n} \left\{ E_{\theta_o} \left(t'(h_i - E_{\theta_o} h)/Y_i, X_i \right) \right\}^2$$

$$= - \frac{1}{2n} V_{\theta_o} \left(t'(h_i - E_{\theta_o} h)/Y_i, X_i \right)$$

It is natural to consider the variable φ^{**}_{in} since it has φ^{*}_{in} as second order expansion. Since B_{in} is non positive, we have $|\varphi^{**}_{in}| \leq 1$.

d.ii) Following the same lines as in parts b) and c) of the proof, it is easily verified that :

$$\boxed{\limsup_{k \to \infty} \; \limsup_{n \to \infty} \; E_{\theta_o} \left| \mathbb{1}_{\Omega_{n,k}} \prod_{i=1}^{n} \varphi^{*}_{in} - \mathbb{1}_{\Omega_{n,k}} \prod_{i=1}^{n} \varphi^{**}_{in} \right| = 0}$$

e) <u>Asymptotic distribution of</u> ξ_n .

i) We have :

$$\prod_{i=1}^{n} \varphi^{**}_{in} = \exp j \left[\sum_{i=1}^{n} A_{in} - j \sum_{i=1}^{n} B_{in} \right]$$

with : $\sum_{i=1}^{n} A_{in} - j \sum_{i=1}^{n} B_{in}$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} E_{\theta_o}\left(t'(h_i - E_{\theta_o} h)/Y_i, X_i\right) + \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \theta'} E_{\theta_o} (t'h_i/Y_i, X_i)$$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} a(Y_i, X_i) + \frac{j}{2n} \sum_{i=1}^{n} V_{\theta_o} (t'(h_i - E_{\theta_o} h)/Y_i, X_i)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} E_{\theta_o}\left(t'(h_i - E_{\theta_o} h)/Y_i, X_i\right) + E_{\theta_o}\left[\frac{\partial}{\partial \theta'} E_{\theta_o} (t'h_i/Y_i, X_i)\right]$$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} a(Y_i, X_i) + \frac{j}{2} E_{\theta_o} V_{\theta_o} (t'(h_i - E_{\theta_o} h)/Y_i, X_i) + o_p(1)$$

where $o_p(1)$ is a negligible term in probability. Since the first three terms converge in distribution, we deduce that the same is true for $\sum_{i=1}^{n} A_{in} - j \sum_{i=1}^{n} B_{in}$ and that the limit of $E_{\theta_o} \prod_{i=1}^{n} \varphi_{in}^{**}$ coincides with the asymptotic characteristic function associated with these terms :

$$\lim_{n\to\infty} E_{\theta_o} \prod_{i=1}^{n} \varphi_{in}^{**} = \exp\left[j\ t'\ m - \frac{t'\ \Sigma\ t}{2}\right]$$

with : $m = E_{\theta_o} E_{\theta_o} \left[(h_i - E_{\theta_o} h)\ /Y_i, X_i)\right]$

$$+ E_{\theta_o}\left[\frac{\partial}{\partial \theta'} E_{\theta_o} (h_i/Y_i, X_i)\right] E_{\theta_o} a(Y_i, X_i) = 0$$

and $\Sigma = E_{\theta_o} V_{\theta_o} \left[(h_i - E_{\theta_o} h)\ /Y_i, X_i)\right]$

$$+ \; V_{\theta_o} \left[E_{\theta_o} (h_i / Y_i , X_i) + E_{\theta_o} \left[\frac{\partial}{\partial \theta'} E_{\theta_o} (h_i / Y_i , X_i) \right] a(Y_i , X_i) \right]$$

$$= \; V_{\theta_o} h_i \; - \; V_{\theta_o} E_{\theta_o} (h_i / Y_i , X_i)$$

$$+ \; V_{\theta_o} \left[E_{\theta_o} (h_i / Y_i , X_i) + E_{\theta_o} \left[\frac{\partial}{\partial \theta'} E_{\theta_o} (h_i / Y_i , X_i) \right] a(Y_i , X_i) \right]$$

ii) The end of the proof is obtained by noting that :

$$\left| E_{\theta_o} \exp(jt'\xi_n) \; - \; E_{\theta_o} \prod_{i=1}^{n} \varphi_{in}^{**} \right|$$

$$\leq \left| E_{\theta_o} \left(\prod_{i=1}^{n} \varphi_{in} \right) \; - \; E_{\theta_o} \left(\mathbb{1}_{\Omega_{n,k}} \prod_{i=1}^{n} \varphi_{in} \right) \right|$$

$$+ \; \left| E_{\theta_o} \left(\mathbb{1}_{\Omega_{n,k}} \prod_{i=1}^{n} \varphi_{in} \right) \; - \; E_{\theta_o} \left(\mathbb{1}_{\Omega_{n,k}} \prod_{i=1}^{n} \varphi_{in}^{*} \right) \right|$$

$$+ \; \left| E_{\theta_o} \left(\mathbb{1}_{\Omega_{n,k}} \prod_{i=1}^{n} \varphi_{in}^{*} \right) \; - \; E_{\theta_o} \left(\mathbb{1}_{\Omega_{n,k}} \prod_{i=1}^{n} \varphi_{in}^{**} \right) \right|$$

$$+ \; \left| E_{\theta_o} \left(\mathbb{1}_{\Omega_{n,k}} \prod_{i=1}^{n} \varphi_{in}^{**} \right) \; - \; E_{\theta_o} \left(\prod_{i=1}^{n} \varphi_{in}^{**} \right) \right|$$

$$\leq \; 2 \, P_{\theta_o} [{}^{c}\Omega_{n,k}] + E_{\theta_o} \left| \mathbb{1}_{\Omega_{n,k}} \prod_{i=1}^{n} \varphi_{in} \; - \; \mathbb{1}_{\Omega_{n,k}} \prod_{i=1}^{n} \varphi_{in}^{*} \right|$$

$$+ \; E_{\theta_o} \left| \mathbb{1}_{\Omega_{n,k}} \prod_{i=1}^{n} \varphi_{in}^{*} \; - \; \mathbb{1}_{\Omega_{n,k}} \prod_{i=1}^{n} \varphi_{in}^{**} \right|$$

and, since the right hand side member tends to zero when n and k tend to infinity, we get :

$$\lim_{n\to\infty} E_{\theta_o} \exp(jt'\xi_n) = \lim_{\xi\to\infty} E_{\theta_o} \prod_{i=1}^{n} \varphi_{in}^{**} = \exp - \frac{t'\,\Sigma\, t}{2} .$$

Q.E.D.

A MODIFIED VERSION OF THE GENERALISED CENTRAL LIMIT THEOREM.

i) Another set of regularity conditions.

The initial set of regularity conditions H1 to H9 may be replaced by H1 , H2, H3, H5, H6 and

H4′ There exists a neighbourhood V of θ_o in which the likelihood function has a derivative with respect to θ and this derivative is continuous at θ_o .

H10 There exists a measurable positive function M from $R^{d_0+d_1+d_2}$ into R such that :

a) the product $||h(Y^*,Y,X)||\ M(Y^*,Y,X)$ is P_{θ_o} integrable,

b) $\left| \frac{\partial}{\partial\theta_j} l(Y^*/Y,X;\theta) \right| \leq M(Y^*,Y,X)$ P_{θ_o} a.s. for any j and for any $\theta \in V$.

In effect condition H4′ , H10 imply H7, H8, H9 .

H4′ , H10 imply H7 .

We have :

$$E_\theta\, [t'\, h(Y^*,Y,X)/Y,X] = t' \int h(y^*,Y,X)\ l(y^*/Y,X;\theta)\ dP_{\theta_o}^{Y^*/Y,X}(y^*)$$

A direct application of the dominated convergence theorem allows to show that the function $\theta \to E_\theta [t'h(Y^*,Y,X)/Y,X]$ is derivable on V and that the derivative is given by :

$$\frac{\partial}{\partial\theta} E_\theta\, [t'h(Y^*,Y,X)/Y,X)] = t' \int h(y^*,Y,X)\ \frac{\partial}{\partial\theta}\ l(y^*/Y,X;\theta)\ dP_{\theta_o}^{Y^*/Y,X}(y^*)$$

In particular the value of this derivative at θ_o is :

$$\left(\frac{\partial}{\partial\theta} E_\theta [t'h(Y^*,Y,X)/Y,X)]\right)_{\theta=\theta_o} = t' E_{\theta_o} \left(h(Y^*,Y,X) \frac{\partial \text{ Log } l}{\partial\theta} [Y^*/Y,X;\theta_o]/Y,X\right)$$

Moreover another application of the dominated convergence theorem and the use of the continuity of $\frac{\partial}{\partial\theta} l(Y^*/Y,X;\theta)$ at θ_o gives the continuity at θ_o of the derivative $\frac{\partial}{\partial\theta} E_\theta [t'h(Y^*,Y,X)/Y,X]$.

H4' , H10 imply H8 :

This is a direct consequence of the inequality :

$$\left|\frac{\partial}{\partial\theta_j} l(Y^*/Y,X;\theta)\right| \leq M(Y^*,Y,X) ,$$

of the integrability of $||h(Y^*,Y,X)||\, M(Y^*,Y,X)$

and of the expression of the derivative :

$$E_{\theta_o} \left(\frac{\partial}{\partial\theta} E_{\theta_o} [t'h(Y^*,Y,X)/Y,X]\right) = t' E_{\theta_o} \left(h(Y^*,Y,X) \frac{\partial l(Y^*,Y,X;\theta_o)}{\partial\theta}\right)$$

H4' , H10 imply H9 :

On the neighbourhood V of θ_o , we have :

$$\left|\left| \frac{\partial}{\partial\theta} E_\theta [t'h(Y^*,Y,X)/Y,X] - \frac{\partial}{\partial\theta} E_{\theta_o} [t'h(Y^*,Y,X)/Y,X]\right|\right|$$

$$\leq E_{\theta_o} \left[||t'h(Y^*,Y,X) [\frac{\partial}{\partial\theta} l(Y^*/Y,X;\theta) - \frac{\partial}{\partial\theta} l(Y^*/Y,X;\theta_o)]\, ||/Y,X\right]$$

$$\leq K\, ||t||\, E_{\theta_o} [||h(Y^*,Y,X)||\; ||M(Y^*,Y,X)|| \,/\, Y,X]$$

where K is a finite number.

Therefore, from H10 a),

$$E_{\theta_o} \left[\sup_{\theta \in V} \left\| \frac{\partial}{\partial \theta} E_{\theta} [t' h(Y^*,Y,X)/Y,X] - \frac{\partial}{\partial \theta} E_{\theta_o} [t' h(Y^*,Y,X)/Y,X] \right\| \right]$$

$$\leq K \, \|t\| \, E_{\theta_o} [\|h(Y^*,Y,X)\| \, \|M(Y^*,Y,X)\|] < +\infty$$

ii) Second form of the asymptotic covariance matrix.

Under the set of conditions H1, H2, H3, H4′, H5, H6, H10, it is possible to give another expression of the asymptotic covariance matrix of ξ_n . This is a consequence of the equality :

$$E_{\theta_o} \left(\frac{\partial}{\partial \theta'} E_{\theta}(h(Y^*,Y,X)/Y,X) \right)_{\theta=\theta_o} = E_{\theta_o} \left[h(Y^*,Y,X) \frac{\partial}{\partial \theta'} \mathrm{Log}\ l(Y^*,Y,X;\theta_o) \right]$$

we have :

$$V_{as}\,\xi = V_{\theta_o} h - V_{\theta_o} E_{\theta_o} (h/Y,X) + V_{\theta_o} \left(E_{\theta_o} (h/Y,X) + E_{\theta_o} \left(h \frac{\partial}{\partial \theta'} \mathrm{Log}\ l(Y^*/Y,X;\theta_o)\right) a(Y,X) \right)$$

$$= E_{\theta_o} V_{\theta_o} (h/Y,X) + V_{\theta_o} \left(E_{\theta_o} (h/Y,X) + E_{\theta_o} \left(h \frac{\partial}{\partial \theta'} \mathrm{Log}\ l(Y^*/Y,X;\theta_o)\right) a(Y,X) \right)$$

where K is a finite number.

Therefore, from H10 a),

$$E_{\theta_o}\left[\sup_{\theta\in V}\left|\left|\frac{\partial}{\partial\theta}E_{\theta}[t'h(Y^*,Y,X)/Y,X] - \frac{\partial}{\partial\theta}E_{\theta_o}[t'h(Y^*,Y,X)/Y,X]\right|\right|\right]$$

$$\leq K\,||t||\,E_{\theta_o}[||h(Y^*,Y,X)||\;||M(Y^*,Y,X)||] < +\infty$$

ii) <u>Second form of the asymptotic covariance matrix.</u>

Under the set of conditions H1, H2, H3, H4', H5, H6, H10, it is possible to give another expression of the asymptotic covariance matrix of ξ_n. This is a consequence of the equality :

$$E_{\theta_o}\left(\frac{\partial}{\partial\theta'}E_{\theta}(h(Y^*,Y,X)/Y,X)\right)_{\theta=\theta_o} = E_{\theta_o}\left[h(Y^*,Y,X)\frac{\partial}{\partial\theta'}\text{Log } l(Y^*,Y,X;\theta_o)\right]$$

we have :

$$V_{as}\xi = V_{\theta_o}h - V_{\theta_o}E_{\theta_o}(h/Y,X) + V_{\theta_o}\left(E_{\theta_o}(h/Y,X)\right.$$

$$\left. + E_{\theta_o}(h\frac{\partial}{\partial\theta'}\text{Log } l(Y^*/Y,X;\theta_o))\;a(Y,X)\right)$$

$$= E_{\theta_o}V_{\theta_o}(h/Y,X) + V_{\theta_o}\left(E_{\theta_o}(h/Y,X)\right.$$

$$\left. + E_{\theta_o}(h\frac{\partial}{\partial\theta'}\text{Log } l(Y^*/Y,X;\theta_o))\;a(Y,X)\right)$$

APPENDIX 2

Strong law of large numbers and consistency of $\tilde{\theta}_n$

---

i) - Strong law of large numbers

- Theorem and regularity assumptions

The notations of appendix 1 and the assumptions H1, H2, H3 are maintained. h is a function, whose values are in $\mathbb{R}^H$, and we denote :

$$h_{\Theta_o} (Y_i^*, Y_i, X_i) = h (Y_i^*, Y_i, X_i) - E_{\Theta_o} h$$

Theorem :

Under assumptions A1 and A2 given below, $\frac{1}{n} \sum_{i=1}^{n} h_{\Theta_o} (Z_{in}, Y_i, X_i)$ converges to 0, P_{Θ_o} almost surely.

A1 : Since the Z_{in} are drawn from $\ell (y^* | y_i, x_i, \hat{\Theta}_n)$ and not from $\ell (y^* | y_i, x_i, \Theta_o)$ we have to make an assumption about the speed of convergence of $\hat{\Theta}_n$ to Θ_o. We do this through iterated logarithm conditions :

$$\begin{cases} P_{\Theta_o} \left[\limsup \dfrac{\sqrt{n} (\hat{\Theta}_n^k - \Theta_o^k)}{\sqrt{\text{LogLogn}}} = \alpha_k \right] = 1 \\ P_{\Theta_o} \left[\liminf \dfrac{\sqrt{n} (\hat{\Theta}_n^k - \Theta_o^k)}{\sqrt{\text{LogLogn}}} = \beta_k \right] = 1 \\ k = 1, \dots, K \end{cases}$$

where $\hat{\Theta}_n^k$ (resp Θ_o^k) are the components of $\hat{\Theta}_n$ (resp Θ_o) and α_k, β_k are real numbers.

A2 : Let $M(\gamma)$ be :

$$M(\gamma) = \sum_{k=1}^{K} \text{Max}\,[(\alpha_k + \gamma)^2, (\beta_k - \gamma)^2]$$

There exists $\gamma > 0$ such that, $\forall\, \varepsilon > 0$:

$$\sum_{n=1}^{\infty} \sqrt{P_{\Theta_o}\, [\tfrac{1}{n} | \sum_{i=1}^{n} h_{\Theta_o}(Y_i^*, Y_i, X_i) | > \varepsilon]\; \eta_n(\gamma)} < +\infty$$

where :

$$\eta_n(\gamma) = \exp\{n\, E_{\Theta_o} \sup_{\|\Theta - \Theta_o\|^2 \leq M(\gamma) \frac{\text{LogLogn}}{n}} |\ell^2(Y_i^* | Y_i, X_i, \Theta) - 1|\}$$

This condition describes the speed of convergence of $\frac{1}{n} \sum_{i=1}^{n} h_{\Theta_o}(Y_i^*, Y_i, X_i)$ to zero :

this speed increases with an average difference between :

$\ell^2(Y_i^* | Y_i, X_i, \hat{\Theta}_n)$ and $\ell^2(Y_i^* | Y_i, X_i, \Theta_o) = 1$.

- Proof of the theorem

First step :

Let us define :

$$A_p = \bigcap_{n\geq p} \left\{ \frac{\sqrt{n}\,(\hat{\theta}_n^k - \theta_o^k)}{\sqrt{\mathrm{LogLog}n}} \leq \alpha_k + \gamma \; ; \; k = 1 \ldots K \right\}$$

$$B_p = \bigcap_{n\geq p} \left\{ \frac{\sqrt{n}\,(\theta_o^k - \hat{\theta}_n^k)}{\sqrt{\mathrm{LogLog}n}} \leq - \beta_k + \gamma \; ; \; k = 1 \ldots K \right\}$$

From assumption A1, we have :

$$\lim_{p=\infty} \uparrow P_{\theta_o} (A_p) = 1 = \lim_{p=\infty} \uparrow P_{\theta_o} (B_p)$$

If we consider :

$$C_p = \bigcap_{n\geq p} \left\{ \frac{n\,||\hat{\theta}_n - \theta_o||^2}{\mathrm{LogLog}n} \leq M(\gamma) \right\}$$

we have $C_p \supset A_p \cap B_p$

and therefore :

$$\lim_{p=\infty} \uparrow P_{\theta_o} (C_p) = 1$$

Second step :

To prove the theorem, we shall show that :

$$\forall \varepsilon > 0 \quad P_{\Theta_o} \left[\bigcup_{n \geq p} \left\{ \frac{1}{n} \left| \sum_{i=1}^{n} h_{\Theta_o} (Z_{in}, Y_i, X_i) \right| > \varepsilon \right\} \right] \xrightarrow[p=\infty]{} 0$$

From the first step, this will be shown if :

$$\forall \varepsilon > 0 \quad P_{\Theta_o} \left[C_p \cap \bigcup_{n \geq p} \left\{ \frac{1}{n} \left| \sum_{i=1}^{n} h_{\Theta_o} (Z_{in}, Y_i, X_i) \right| > \varepsilon \right\} \right] \xrightarrow[p=\infty]{} 0$$

or if :

$$\forall \varepsilon > 0 \quad \sum_{n=p}^{\infty} P_{\Theta_o} \left[C_p \cap \left\{ \frac{1}{n} \left| \sum_{i=1}^{n} h_{\Theta_o} (Z_{in}, Y_i, X_i) \right| > \varepsilon \right\} \right] \xrightarrow[p=\infty]{} 0$$

or, since the C_n are increasing, if :

$$\forall \varepsilon > 0 \quad \sum_{n=1}^{\infty} P_{\Theta_o} \left[C_n \cap \left\{ \frac{1}{n} \left| \sum_{i=1}^{n} h_{\Theta_o} (Z_{in}, Y_i, X_i) \right| > \varepsilon \right\} \right] < + \infty$$

Third step :

$\mathbb{1}_{C_n}$ depends on $X_1, \ldots, X_n, Y_1^*, \ldots, Y_n^*$ through $\hat{\Theta}_n$. Thus, it only depends on $X_1, \ldots, X_n, Y_1, \ldots, Y_n$ and :

$$P_{\Theta_o} \left[C_n \cap \left\{ \frac{1}{n} \left| \sum_{i=1}^{n} h_{\Theta_o} (Z_{in}, Y_i, X_i) \right| > \varepsilon \right\} \right]$$

$$= E_{\Theta_o} \left[\mathbb{1}_{C_n} \; P_{\Theta_o} \left[\frac{1}{n} \left| \sum_{i=1}^{n} h_{\Theta_o} (Z_{in}, Y_i, X_i) \right| > \varepsilon \;\middle|\; (Y_i, X_i)_{1 \leq i \leq n} \right] \right]$$

But :

$$P_{\Theta_o}\ [\ \frac{1}{n}\ |\ \sum_{i=1}^{n} h_{\Theta_o}\ (Z_{in},\ Y_i,\ X_i)\ |\ >\ \varepsilon\ |\ (Y_i,\ X_i)_{1\leq i\leq n} = (y_i,\ x_i)_{1\leq i\leq n}]$$

$$= \int \mathbb{1}_{\frac{1}{n}\ |\ \sum_{i=1}^{n} h_{\Theta_o}\ (y_i^*,\ y_i,\ x_i)\ |\ >\ \varepsilon}\ \bigotimes_{i=1}^{n} P_{\hat{\Theta}_n}^{Y^*\ |\ (Y_i,\ X_i)\ =\ (y_i,\ x_i)}\ (dy_i^*)$$

with :

$$P_{\hat{\Theta}_n}^{Y^*\ |\ (Y_i,\ X_i)\ =\ (y_i,\ x_i)}\ (dy_i^*) = \ell\ (y_i^*\ |\ y_i,\ x_i,\ \hat{\Theta}_n)\ \ P_{\Theta_o}^{Y^*\ |\ (Y_i,\ X_i)\ =\ (y_i,\ x_i)}(dy_i^*)$$

Thus :

$$P_{\Theta_o}\ [\ \frac{1}{n}\ |\ \sum_{i=1}^{n} h_{\Theta_o}\ (z_{in},\ Y_i,\ X_i)\ |\ >\ \varepsilon\ |\ (Y_i,\ X_i)_{1\leq i\leq n} = (y_i,\ x_i)_{1\leq i\leq n}\ ]$$

$$= E_{\Theta_o}\ [\ \mathbb{1}_{\frac{1}{n}\ |\ \sum_{i=1}^{n} h_{\Theta_o}\ (Y_i^*,\ Y_i,\ X_i)\ |\ >\ \varepsilon}\ \prod_{i=1}^{n} \ell\ (Y_i^*|Y_i,\ X_i,\ \hat{\Theta}_n)\ |\ (Y_i,\ X_i)_{1\leq i\leq n} = (y_i,\ x_i)_{1\leq i\leq n}\ ]$$

Therefore, by the Cauchy-Schwarz inequality :

$$P_{\Theta_o}\ [\ C_n\ \cap \{\ \frac{1}{n}\ |\ \sum_{i=1}^{n} h_{\Theta_o}\ (Z_{in},\ Y_i,\ X_i)\ |\ >\ \varepsilon\}]$$

$$= E_{\Theta_o}\ [\ \mathbb{1}_{C_n}\ \mathbb{1}_{\frac{1}{n}\ |\ \sum_{i=1}^{n} h_{\Theta_o}\ (Y_i^*,\ Y_i,\ X_i)\ |\ >\ \varepsilon}\ \prod_{i=1}^{n} \ell\ (Y_i^*|Y_i,\ X_i,\ \hat{\Theta}_n)]$$

$$\leq \sqrt{P_{\Theta_o}\left[\frac{1}{n}\left|\sum_{i=1}^{n} h_{\Theta_o}(Y_i^*, Y_i, X_i)\right| > \varepsilon\right]}\ \sqrt{E_{\Theta_o}\left[\prod_{i=1}^{n} 1_{C_n}\, \ell^2(Y_i^*|Y_i, X_i, \hat{\Theta}_n)\right]}$$

Thus, in order to prove the theorem, it is sufficient to show that, $\forall\ \varepsilon > 0$

$$\sum_{n=1}^{\infty} \sqrt{P_{\Theta_o}\left[\frac{1}{n}\left|\sum_{i=1}^{n} h_{\Theta_o}(Y_i^*, Y_i, X_i)\right| > \varepsilon\right]}\ \sqrt{E_{\Theta_o}\left[\prod_{i=1}^{n} 1_{C_n}\, \ell^2(Y_i^*|Y_i, X_i, \hat{\Theta}_n)\right]} < +\infty$$

<u>Fourth step</u> :

By definition of C_n :

$$\left| 1_{C_n}\, \ell^2 (Y_i^* \mid Y_i, X_i, \hat{\Theta}_n) - 1_{C_n} \right|$$

$$\leq \sup_{||\Theta - \Theta_o||^2 \leq \frac{M(\gamma)\,\mathrm{LogLog}n}{n}} \left| \ell^2 (Y_i^* \mid Y_i, X_i, \Theta) - 1 \right| = \alpha_{in}$$

Since the random variables α_{in}, $i = 1 \ldots n$, are i.i.d., we have from the lemma of appendix 1 :

$$E_{\Theta_o}\left[\prod_{i=1}^{n} 1_{C_n}\, \ell^2 (Y_i^*|Y_i, X_i, \hat{\Theta}_n)\right]$$

$$\leq E_{\Theta_o}\left[\prod_{i=1}^{n} 1_{C_n}\right] + \exp\left[\sum_{i=1}^{n} E_{\Theta_o} \alpha_{in}\right] - 1$$

$$\leq \exp\left[\sum_{i=1}^{n} E_{\Theta_o} \alpha_{in}\right]$$

$$= \exp \left[n \; E_{\Theta_o} \sup_{\| \Theta - \Theta_o \|^2 \leq \frac{M(\gamma) \; \mathrm{LogLog} n}{n}} \left| \ell^2 (Y_i^* | Y_i, X_i, \Theta) - 1 \right| \right] = \eta_n (\gamma)$$

which achieves the proof with the assumption A2 and the third step.□

In the sequel it will be also useful to consider the uniform strong convergence of functions such as :

$$\tilde{h}_{\theta o} (Y_i^*, Y_i, X_i ; \theta) = \tilde{h} (Y_i^*, Y_i, X_i ; \theta) - E_{\theta o} \tilde{h}$$

<u>Corollary</u>

If A_1 is satisfied and if A_2 is replaced by a similar condition A_2' obtained from A_2 by replacing

$$\left| \sum_{i=1}^{n} h_{\theta_o} (Y_i^*, Y_i ; X_i) \right| \quad \text{by}$$

$$\sup_{\theta \in \Theta} \left| \sum_{i=1}^{n} \tilde{h}_{\theta_o} (Y_i^*, Y_i, X_i ; \theta) \right| \quad , \text{then } \frac{1}{n} \sum_{i=1}^{n} \tilde{h}_{\theta_o} (Z_{in}, Y_i, X_i)$$

converges P_{θ_o} almost surely to 0 , uniformly in $\theta \in \Theta$

<u>Proof</u>: it is the same as the previous one. □

Typically $\eta_n(\gamma)$ will be equivalent to $\exp \; \alpha_1 \sqrt{n \; \mathrm{Log} \; \mathrm{Log} \; n}$ $(\alpha_1 > o)$; on the other hand, when $\tilde{h}(Y_i^*, Y_i, X_i ; \theta)$ is a p.d.f. of Y_i^* given X_i which is gaussian or derived from a gaussian distribution the probability appearing in A_2 or A_2' will be equivalent to

$\alpha_2 \, n^{-\alpha_3} \exp[- \alpha_4 n] \; [\alpha_2 > o, \alpha_3 \geq o, \alpha_4 > o]$; in such cases

A2 , or A_2' , is satisfied.

ii) Consistency of $\tilde{\theta}_n$

a) When the latent model is the linear model

$$y_i^* = x_i\, b + u_i$$

where u_i is distributed, conditionally or x_i , as $N(0,\sigma^2)$, the vector $\tilde{\theta}_n'$ is $(\tilde{b}_n', \tilde{\sigma}_n^2)$ where

$$\tilde{b}_n = (\sum_{i=1}^{n} x_i' x_i)^{-1} \sum_{i=1}^{n} x_i' z_{in}$$

$$\tilde{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} (z_{in} - x_i \tilde{b}_n)^2$$

In this case $\tilde{\theta}_n$ can be explicitly expressed in terms of the z_{in} and x_i ; therefore the weak consistency of $\tilde{\theta}_n$ is a straightforward consequence of corollary 2.6. Note that this weak consistency is sufficient for the proof of theorem 3.4.

b) If the latent model is the non linear regression model

$$y_i^* = k(x_i, b) + u_i$$

where the u_i are, conditionally on x_i , distributed as $N(0,\sigma^2)$, the log-likelihood function is

$$L_n = -\frac{n}{2} \text{Log } 2\pi - \frac{n}{2} \text{Log } \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} [y_i^* - k(x_i, b)]^2$$

$$= -\frac{n}{2} \text{Log } 2\pi - \frac{n}{2} \text{Log } \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} [y_i^* - k(x_i, b_o)]^2$$

$$+ \frac{1}{\sigma^2} \sum_{i=1}^{n} [y_i^* - k(x_i,b_o)][k(x_i,b_o) - k(x_i,b)]$$

$$- \frac{1}{2\sigma^2} \sum_{i=1}^{n} [k(x_i,b_o) - k(x_i,b)]^2$$

Therefore : $\frac{L_n}{n}$ converges P_{Θ_o} a.s., and uniformly in Θ , $[\Theta' = (b',\sigma^2)]$

to $-\frac{1}{2}\,\text{Log}\,2\,\pi - \frac{1}{2}\,\text{Log}\,\sigma^2 - \frac{1}{2\sigma^2}\,E_{\Theta_o}[Y_i^*-k(X_i,b_o)]^2 - \frac{1}{2\sigma^2}\,E_{\Theta_o}[k(X,b_o)-k(X,b)]^2$

$$= -\frac{1}{2}\,\text{Log}\,2\,\pi - \frac{1}{2}\,\text{Log}\,\sigma^2 - \frac{\sigma_o^2}{2\sigma^2} - \frac{1}{2\sigma^2}\,E_{\Theta_o}\,[k(X,b_o) - k(X,b)]^2$$

provided that

$\frac{1}{n}\sum_{i=1}^{n}[y_i^*-k(x_i,b_o)]^2$ converges P_{Θ_o} p.s. to σ_o^2

$\frac{1}{n}\sum_{i=1}^{n}[y_i^*-k(x_i,b_o)][k(x_i,b_o)-k(x_i,b)]$ converges P_{Θ_o} p.s. to 0 uniformly in b

$\frac{1}{n}\sum_{i=1}^{n}[k(x_i,b_o) - k(x_i,b)]^2$ converges P_{Θ_o} p.s. to $E_{\Theta_o}[k(X,b_o)-k(X,b)]^2$ uniformly in b

In fact, using Cauchy-Schwarz inequality, it is readily seen that the second condition is implied by the two other conditions. In particular we do not need a uniform convergence in which the y_i^* are involved.

As a consequence, when considering the strong consistency of $\tilde{\Theta}_n$, i.e. when the y_i^* are replaced by the z_{in}, we only need the version of the previous theorem without uniformity in order to obtain the uniform strong convergence of the objective function to

$$-\frac{1}{2}\,\text{Log}\,2\,\pi - \frac{1}{2}\,\text{Log}\,\sigma^2 - \frac{\sigma_o^2}{2\sigma^2} - \frac{1}{2\sigma^2}\,E_{\Theta_o}[k(X,b_o) - k(X,b)]^2 \quad .$$

Assuming the asymptotic identifiability of b from $k(X,b)$, this limit function has a unique maximum in $\Theta_o^*[\Theta_o' = (b_o, \sigma_o^2)]$, and the strong consistency of $\tilde{\Theta}_n$ is obtained by standard arguments [see e.g. Jennrich, (1969)].

c) In the general case $\tilde{\Theta}_n$ is obtained by maximising

$$\frac{1}{n}\sum_{i=1}^{n}\text{Log}\,\ell^*(z_{in}/x_i;\Theta) \;.$$

Under the assumptions of the previous theorem for the uniform strong convergence, this objective function strongly converges uniformly in θ , to

$$E_{\theta_o} \operatorname{Log} l^*(Y^*/X;\theta)$$

This limit objective function is the same as the limit objective function in the latent model which will be typically assumed to have a unique maximum in θ_o . The strong consistency of $\tilde{\theta}_n$ follows.

Appendix 3

Asymptotic properties of the test statistic ξ_n

i) We assume that are satisfied the regularity conditions allowing the application of the generalised central limit theorem and of the strong law of large numbers of appendices 1 and 2, for the true value $\theta_o = \binom{0}{\beta_o}$ of the parameter. Under these conditions, we have :

$$\operatorname*{plim}_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \text{Log } \ell^* (z_{in}/x_i;0;\beta_o)}{\partial\beta \, \partial\beta'} = -I_{\beta\beta}$$

and the estimator $\tilde{\beta}_{on}$ of β defined as the solution of the equations :

$$\sum_{i=1}^{n} \frac{\partial \text{ Log } \ell^* (z_{in}/x_i;0;\tilde{\beta}_{on})}{\partial\beta} = 0$$

is such that :

$$\sqrt{n}\,(\tilde{\beta}_{on} - \beta_o) = \frac{1}{\sqrt{n}} \; I^{-1}_{\beta\beta} \sum_{i=1}^{n} \frac{\partial \text{ Log } \ell^* (z_{in}/x_i;0;\beta_o)}{\partial\beta} + o_p(1)$$

ii) Let us now consider the asymptotic expansion of the statistic ξ_n. We get :

$$\xi_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \text{ Log } \ell^* (z_{in}/x_i;0;\tilde{\beta}_{on})}{\partial\alpha}$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \text{ Log } \ell^* (z_{in}/x_i;0;\beta_o)}{\partial\alpha}$$

$$+ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \text{ Log } \ell^* (z_{in}/x_i;0;\beta_o)}{\partial\alpha \, \partial\beta'} \; \sqrt{u}\,(\tilde{\beta}_{on} - \beta_o)$$

$$+ o_p(1)$$

Since the empirical mean $- \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \text{ Log } \ell^* (z_{in}/x_i;0;\beta_o)}{\partial\alpha \, \partial\beta'}$

converges to the block :

$$I_{\alpha\beta} = \underset{\Theta_o}{E} \left[- \frac{\partial^2 \text{Log } \ell^* (Y^* / X ; 0 ; \beta_o)}{\partial\alpha \, \partial\beta'} \right]$$

of the latent information matrix (see Corollary 2.6), we also have :

$$\xi_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \text{ Log } \ell^* (z_{in}/x_i ; 0 ; \beta_o)}{\partial\alpha}$$

$$- I_{\alpha\beta} \sqrt{n} \, (\tilde{\beta}_{on} - \beta_o) + o_p(1)$$

Replacing in the expression of ξ_n, $\sqrt{n} \, (\tilde{\beta}_{on} - \beta_o)$ by its asymptotic expansion, we get :

$$\xi_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \text{ Log } \ell^* (z_{in}/x_i ; 0 ; \beta_o)}{\partial\alpha}$$

$$-I_{\alpha\beta} \, I_{\beta\beta}^{-1} \, \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \text{ Log } \ell^* (z_{in}/x_i ; 0 ; \beta_o)}{\partial\beta} + o_p(1)$$

Therefore, the asymptotic normality of ξ_n is directly obtained by applying the generalised central limit theorem with :

$$h(z_{in}, y_i, x_i) = \frac{\partial \text{ Log } \ell^* (z_{in}/x_i ; 0 ; \beta_o)}{\partial\alpha}$$

$$- I_{\alpha\beta} \, I_{\beta\beta}^{-1} \frac{\partial \text{ Log } \ell^* (z_{in}/x_i ; 0 ; \beta_o)}{\partial\beta}$$

iii) ξ_n is asymptotically zero mean, since :

$$\underset{\Theta_o}{E} \; h(Y^*, Y, X)$$

$$= \underset{\Theta_o}{E} \left[\frac{\partial \text{ Log } \ell^* (Y^*/X ; 0 , \beta_o)}{\partial\alpha} \right] - I_{\alpha\beta} \, I_{\beta\beta}^{-1} \underset{\Theta_o}{E} \left[\frac{\partial \text{ Log } \ell^* (Y^*/X ; 0 ; \beta_o)}{\partial\beta} \right]$$

is equal to zero as a combination of the expectation of the score vector.

iv) The asymptotic covariance matrix of ξ_n is :

$$V_{as}\ \xi_n = V_{\theta_o}\ h - V_{\theta_o}\ E_{\theta_o}\ (h/Y,X)$$

$$+ V_{\theta_o}\ \{ E_{\theta_o}\ (h/Y,X) + E_{\theta_o}\ [h \frac{\partial}{\partial \beta'}\ \text{Log}\ \ell\ (Y^*/Y,X;\theta_o)]\ a(Y,X)\}$$

with $a(Y,X) = J^{-1}_{\beta\beta}\ \dfrac{\partial\ \text{Log}\ \ell\ (Y/X;0,\beta_o)}{\partial \beta}$

since the estimator $\hat{\theta}_{on}$ used for the simulations is the constrained maximum likelihood estimator of θ.

It is possible to express $V_{as}\ \xi_n$ in terms of the latent and observable information matrices I and J by using the following equalities :

(1) $$\frac{\partial\ \text{Log}\ \ell\ (Y/X;\theta)}{\partial \theta} = E_{\theta}\ [\ \frac{\partial\ \text{Log}\ \ell^*\ (Y^*/X;\theta)}{\partial \theta}\ /Y,X\ ]$$

(2) $$\frac{\partial\ \text{Log}\ \ell^*\ (Y^*/X;\theta)}{\partial \theta} = \frac{\partial\ \text{Log}\ \ell\ (Y/X;\theta)}{\partial \theta} + \frac{\partial\ \text{Log}\ \ell\ (Y^*/Y,X;\theta)}{\partial \theta}$$

v) <u>Expression of $E_{\theta_o}\ [h \frac{\partial}{\partial \beta'}\ \text{Log}\ \ell\ (Y^*/Y,X;\theta_o)]$</u>

We have :

$$h \frac{\partial}{\partial \beta'}\ \text{Log}\ \ell\ (Y^*/Y,X;\theta_o)$$

$$= \frac{\partial\ \text{Log}\ \ell^*(Y^*/X;\theta_o)}{\partial \alpha}\ \frac{\partial\ \text{Log}\ \ell\ (Y^*/Y,X;\theta_o)}{\partial \beta'}$$

$$-I_{\alpha\beta} \quad I_{\beta\beta}^{-1} \quad \frac{\partial \text{ Log } \ell^* \ (Y^*/X;\theta_o)}{\partial\beta} \qquad \frac{\partial \text{ Log } \ell \ (Y^*/Y,X;\theta_o)}{\partial\beta'}$$

Then, using (2), we get :

$$\underset{\theta_o}{E} \quad [\frac{\partial \text{ Log } \ell^* \ (Y^*/X;\theta_o)}{\partial\alpha} \qquad \frac{\partial \text{ Log } \ell \ (Y^*/Y,X;\theta_o)}{\partial\beta'}]$$

$$= \underset{\theta_o}{E} \quad [\frac{\partial \text{ Log } \ell^* \ (Y^*/X;\theta_o)}{\partial\alpha} \qquad \frac{\partial \text{ Log } \ell^* \ (Y^*/X;\theta_o)}{\partial\beta'}]$$

$$- \underset{\theta_o}{E} \quad [\frac{\partial \text{ Log } \ell^* \ (Y^*/X;\theta_o)}{\partial\alpha} \qquad \frac{\partial \text{ Log } \ell \ (Y/X;\theta_o)}{\alpha\beta'}]$$

$$= I_{\alpha\beta} - J_{\alpha\beta}$$

since $$\underset{\theta_o}{E} \quad [\frac{\partial \text{ Log } \ell^* \ (Y^*/X;\theta_o)}{\partial\alpha} \qquad \frac{\partial \text{ Log } \ell \ (Y/X;\theta_o)}{\partial\beta'}]$$

$$= \underset{\theta_o}{E} \quad \{ \underset{\theta_o}{E} [\frac{\partial \text{ Log } \ell^* \ (Y^*/X;\theta_o)}{\partial\alpha} /Y,X] \qquad \frac{\partial \text{ Log } \ell \ (Y/X;\theta_o)}{\partial\beta'} \}$$

$$= \underset{\theta_o}{E} \quad [\frac{\partial \text{ Log } \ell \ (Y/X;\theta_o)}{\partial\alpha} \qquad \frac{\partial \text{ Log } \ell \ (Y/X;\theta_o)}{\partial\beta'}] \qquad \text{from (1).}$$

Therefore, we deduce :

$$\underset{\theta_o}{E} \quad [h \frac{\partial}{\partial\beta'} \text{ Log } \ell \ (Y^*/Y,X;\theta_o)]$$

$$= I_{\alpha\beta} - J_{\alpha\beta} - I_{\alpha\beta} \ I_{\beta\beta}^{-1} \ [I_{\beta\beta} - J_{\beta\beta}]$$

$$= - J_{\alpha\beta} + I_{\alpha\beta} \ I_{\beta\beta}^{-1} \ J_{\beta\beta}$$

vi) Expression of : $V_{\Theta_o} \{ E_{\Theta_o} (h/Y,X) + E_{\Theta_o} [h \frac{\partial}{\partial \beta'} \text{Log}\, \ell\, (Y^*/Y,X;\theta_o)]\, a(Y,X) \}$

From (1), the conditional expectation $E_{\Theta_o} (h/Y,X)$ is equal to :

$$E_{\Theta_o} (h/Y,X) = \frac{\partial \text{ Log } \ell \; (Y/X;\theta_o)}{\partial \alpha} - I_{\alpha\beta}\, I_{\beta\beta}^{-1} \frac{\partial \text{ Log } \ell \; (Y/X;\theta_o)}{\partial \beta}$$

Then, we have to determine the covariance matrix of :

$$\frac{\partial \text{ Log } \ell \; (Y/X;\theta_o)}{\partial \alpha} - I_{\alpha\beta}\, I_{\beta\beta}^{-1} \frac{\partial \text{ Log } \ell \; (Y/X;\theta_o)}{\partial \beta}$$

$$+ (-J_{\alpha\beta} + I_{\alpha\beta}\, I_{\beta\beta}^{-1}\, J_{\beta\beta})\, J_{\beta\beta}^{-1} \frac{\partial \text{ Log } \ell \; (Y/X;\theta_o)}{\partial \beta}$$

$$= \frac{\partial \text{ Log } \ell \; (Y/X;\theta_o)}{\partial \alpha} - J_{\alpha\beta}\, J_{\beta\beta}^{-1} \frac{\partial \text{ Log } \ell \; (Y/X;\theta_o)}{\partial \beta}$$

This covariance matrix is equal to :

$$J_{\alpha\alpha} - J_{\alpha\beta}\, J_{\beta\beta}^{-1}\, J_{\beta\alpha}$$

vii) Expression of $V_{as}\, \xi_n$

$$V_{as}\; \xi_n = V_{\Theta_o}\, h - V_{\Theta_o}\, E_{\Theta_o} (h/Y,X) + J_{\alpha\alpha} - J_{\alpha\beta}\, J_{\beta\beta}^{-1}\, J_{\beta\alpha}$$

$$= V_{\Theta_o} \left[\frac{\partial \text{ Log } \ell^* \; (Y^*/X;\theta_o)}{\partial \alpha} - I_{\alpha\beta}\, I_{\beta\beta}^{-1} \frac{\partial \text{ Log } \ell^* (Y^*/X;\theta_o)}{\partial \beta} \right]$$

$$- V_{\Theta_o} \left[\frac{\partial \text{ Log } \ell \; (Y/X;\theta_o)}{\partial \alpha} - I_{\alpha\beta}\, I_{\beta\beta}^{-1} \frac{\partial \text{ Log } \ell \; (Y/X;\theta_o)}{\partial \beta} \right.$$

$$+ J_{\alpha\alpha} - J_{\alpha\beta}\, J_{\beta\beta}^{-1}\, J_{\beta\alpha}$$

$$= I_{\alpha\alpha} - I_{\alpha\beta}\, I_{\beta\beta}^{-1}\, I_{\beta\alpha}$$

$$- \{ J_{\alpha\alpha} - I_{\alpha\beta}\, I_{\beta\beta}^{-1}\, J_{\beta\alpha} - J_{\alpha\beta}\, I_{\beta\beta}^{-1}\, I_{\beta\alpha} + I_{\alpha\beta}\, I_{\beta\beta}^{-1}\, J_{\beta\beta}\, I_{\beta\beta}^{-1}\, I_{\beta\alpha} \}$$

$$+ J_{\alpha\alpha} - J_{\alpha\beta}\, J_{\beta\beta}^{-1}\, J_{\beta\alpha}$$

$$= I_{\alpha\alpha} - I_{\alpha\beta}\, I_{\beta\beta}^{-1}\, I_{\beta\alpha} + I_{\alpha\beta}\, I_{\beta\beta}^{-1}\, J_{\beta\alpha} + J_{\alpha\beta}\, I_{\beta\beta}^{-1}\, I_{\beta\alpha}$$

$$- I_{\alpha\beta}\, I_{\beta\beta}^{-1}\, J_{\beta\beta}\, I_{\beta\beta}^{-1}\, I_{\beta\alpha} - J_{\alpha\beta}\, J_{\beta\beta}^{-1}\, J_{\beta\alpha}$$

This expression is equivalent to :

$$V_{as}\, \xi_n = I_{\alpha\alpha} - I_{\alpha\beta}\, I_{\beta\beta}^{-1}\, I_{\beta\alpha}$$

$$- (I_{\alpha\beta}\, I_{\beta\beta}^{-1} - J_{\alpha\beta}\, J_{\beta\beta}^{-1})\; J_{\beta\beta}\; (I_{\beta\beta}^{-1}\, I_{\beta\alpha} - J_{\beta\beta}^{-1}\, J_{\beta\alpha})$$

Appendix 4

Comparison of the asymptotic covariance matrices : $V_{as}\ \xi_n$ and $V_{as}\ \hat{\xi}_n$.

We first consider the case in which the matrix $I_{\beta\beta} - J_{\beta\beta}$ is positive and we have to prove that :

$$V_{as}\ \xi_n = f(I,J) = I_{\alpha\alpha} - I_{\alpha\beta}\ I_{\beta\beta}^{-1}\ I_{\beta\alpha}$$

$$- (I_{\alpha\beta}\ I_{\beta\beta}^{-1} - J_{\alpha\beta}\ J_{\beta\beta}^{-1})\ J_{\beta\beta}(I_{\beta\beta}^{-1}\ I_{\beta\alpha} - J_{\beta\beta}^{-1}\ J_{\beta\alpha})$$

is greater than :

$$\frac{V_{as}\ \hat{\xi}_n}{\lambda} = \frac{I_{\alpha\alpha} - I_{\alpha\beta}\ I_{\beta\beta}^{-1} I_{\beta\alpha}}{\lambda} = \hat{f}(I,J)$$

where λ is the greatest eigenvalue of $J_{\beta\beta}^{-1/2}\ I_{\beta\beta}\ J_{\beta\beta}^{-1/2}$

i) Orthogonalisation

We can first remark that it is possible to assume that $I_{\alpha\beta} = o$. In effect, if it is not the case, we can introduce the invertible matrix :

$$P = \left| \begin{array}{c:c} I_{K_1} & -I_{\alpha\beta}\ I_{\beta\beta}^{-1} \\ \hdashline 0 & I_{K_2} \end{array} \right|$$

and it is easily seen that the condition :

$$f(I,J) >> \hat{f}(I,J)$$

is equivalent to :

$$f(\tilde{I},\tilde{J}) >> \hat{f}(\tilde{I},\tilde{J})$$

with $\tilde{I} = P\ I\ P'$ and $\tilde{J} = P\ J\ P'$. The possibility of choosing $I_{\alpha\beta} = o$ is a consequence of the form of $\tilde{I}$:

$$\tilde{I} = \left|\begin{array}{c|c} I_{\alpha\alpha} - I_{\alpha\beta} I_{\beta\beta}^{-1} I_{\beta\alpha} & 0 \\ \hline 0 & I_{\beta\beta} \end{array}\right|$$

ii) Under the condition $I_{\alpha\beta} = o$, the inequality to be proved is :

$$I_{\alpha\alpha} - J_{\alpha\beta} J_{\beta\beta}^{-1} J_{\beta\alpha} >> \frac{I_{\alpha\alpha}}{\lambda}$$

where λ is the maximum eigenvalue of $J_{\beta\beta}^{-1/2} I_{\beta\beta} J_{\beta\beta}^{-1/2}$.

This inequality is equivalent to :

$$(1) \quad J_{\alpha\beta} J_{\beta\beta}^{-1} J_{\beta\alpha} << I_{\alpha\alpha}(1 - \frac{1}{\lambda}) = I_{\alpha\alpha} \frac{1}{1+\mu}$$

where μ is the inverse of the maximum eigenvalue of $J_{\beta\beta}^{-1/2} I_{\beta\beta} J_{\beta\beta}^{-1/2} - I_{K_2}$.

iii) To prove this inequality, we only know that the latent information matrix I is greater than the observable one and that they are positive :

$$I >> J >> 0$$

$$\Leftrightarrow \begin{bmatrix} I_{\alpha\alpha} - J_{\alpha\alpha} & -J_{\alpha\beta} \\ -J_{\beta\alpha} & I_{\beta\beta} - J_{\beta\beta} \end{bmatrix} >> 0 \quad \text{and} \quad \begin{bmatrix} J_{\alpha\alpha} & J_{\alpha\beta} \\ J_{\beta\alpha} & J_{\beta\beta} \end{bmatrix} >> 0$$

$$\Leftrightarrow \begin{cases} J_{\beta\beta} >> 0 \text{ , } J_{\alpha\alpha} - J_{\alpha\beta} J_{\beta\beta}^{-1} J_{\beta\alpha} >> o \\ I_{\beta\beta} - J_{\beta\beta} >> o \text{ , } I_{\alpha\alpha} - J_{\alpha\alpha} - J_{\alpha\beta}(I_{\beta\beta} - J_{\beta\beta})^{-1} J_{\beta\alpha} >> o \end{cases}$$

iv) Let us denote by $K_{\alpha\beta}$ the matrix $J_{\alpha\beta} J_{\beta\beta}^{-1/2}$ and by $K_{\beta\alpha}$ the transposed matrix $K'_{\alpha\beta}$. With these new notations, two known inequalities are :

$$\begin{cases} J_{\alpha\alpha} - K_{\alpha\beta} K_{\beta\alpha} >> 0 \\ I_{\alpha\alpha} - J_{\alpha\alpha} - K_{\alpha\beta}[J_{\beta\beta}^{-1/2} I_{\beta\beta} J_{\beta\beta}^{-1/2} - I_{K_2}]^{-1} K_{\beta\alpha} >> o \end{cases}$$

and the required inequality becomes :

$$(2) \quad K_{\alpha\beta}\, K_{\beta\alpha} << I_{\alpha\alpha} \frac{1}{1+\mu}$$

v) Let us now consider : a spectral decomposition of the matrix $[J_{\beta\beta}^{-1/2}\, I_{\beta\beta}\, J_{\beta\beta}^{-1/2} - I_{K_2}]^{-1}$; we have :

$[J_{\beta\beta}^{-1/2}\, I_{\beta\beta}\, J_{\beta\beta}^{-1/2} - I_{K_2}]^{-1} = Q\, \Lambda\, Q'$ where Q is an orthogonal matrix with size K_2 and Λ a diagonal matrix with elements μ_k , $k = 1 \dots K_2$.
If we denote by H_k , $k = 1 \dots K_2$ the K_2 column vectors of $K_{\alpha\beta}\, Q$, the known inequalities become :

$$(3) \begin{cases} \sum_{k=1}^{K_2} H_k H_k' << J_{\alpha\alpha} \\ \sum_{k=1}^{K_2} \mu_k H_k H_k' << I_{\alpha\alpha} - J_{\alpha\alpha} \end{cases}$$

and the required inequality is :

$$(4) \quad \sum_{k=1}^{K_2} H_k H_k' << I_{\alpha\alpha} \frac{1}{1+\mu} \quad \text{with} \quad \mu = \underset{k}{\text{Min}}\, \mu_k \; .$$

vi) The proof is completed by noting that (3) implies :

$$\sum_{k=1}^{K_2} (1 + \mu_k)\, H_k H_k' << I_{\alpha\alpha}$$

and also :

$$(1+\mu) \sum_{k=1}^{K_2} H_k H_k' << I_{\alpha\alpha}$$

$$\Leftrightarrow \sum_{k=1}^{K_2} H_k H_k' << I_{\alpha\alpha} \frac{1}{1+\mu}$$

vii) Finally, the result is easily extended to the case in which $I_{\beta\beta} - J_{\beta\beta}$ is only non negative. In such a case, we can introduce the matrix $J^{\varepsilon} = \frac{1}{1+\varepsilon} J$, where ε is a real positive number.
Since J is positive, J^{ε} is positive and the same is true for :

$$I_{\beta\beta} - J^{\varepsilon}_{\beta\beta} = I_{\beta\beta} - J_{\beta\beta} + \frac{\varepsilon}{1+\varepsilon} J_{\beta\beta} >> \frac{\varepsilon}{1+\varepsilon} J_{\beta\beta}$$

Then the previous result can be applied to the pair (I, J^{ε}) and leads to the inequality

$$J^{\varepsilon}_{\alpha\beta} (J^{\varepsilon}_{\beta\beta})^{-1} J^{\varepsilon}_{\beta\alpha} << (1 - \frac{1}{\lambda^{\varepsilon}}) I_{\alpha\alpha}$$

where λ^{ε}, maximum eigenvalue of $(J^{\varepsilon}_{\beta\beta})^{-1/2} I_{\beta\beta} (J^{\varepsilon}_{\beta\beta})^{-1/2}$, is equal to :

$$\lambda^{\varepsilon} = (1+\varepsilon)\ \lambda$$

In terms of J and I the inequality can be written :

$$\frac{1}{1+\varepsilon} J_{\alpha\beta} J^{-1}_{\beta\beta} J_{\beta\alpha} << (1 - \frac{1}{\lambda(1+\varepsilon)}) I_{\alpha\alpha}$$

and we only have to let ε go to zero to obtain the required inequality :

$$J_{\alpha\beta} J^{-1}_{\beta\beta} J_{\beta\beta} << (1 - \frac{1}{\lambda}) I_{\alpha\alpha}$$

Appendix 5

Asymptotic properties of the test statistic $\bar{\xi}_n$

Let $\tilde{\alpha}_n$ be the unconstrained second stage estimator of α , based on the z_{in} obtained from the unconstrained M.L. estimator. From the result of subsection 5.d, we know that $\bar{\xi}_n$ is asymptotically equivalent to $\sqrt{n}\,[I^{\alpha\alpha}(\Theta_o)]^{-1}\tilde{\alpha}_n$. Moreover the asymptotic distribution of $\sqrt{n}\,\tilde{\alpha}_n$ is obtained from theorem 3.4 : it is the zero-mean normal distribution whose covariance matrix is (dropping Θ_o) :

$$I^{\alpha\alpha} - (I^{\alpha\alpha} \vdots I^{\alpha\beta})\, J (I^{\alpha\alpha} \vdots I^{\alpha\beta})' + J^{\alpha\alpha}$$

Therefore $\bar{\xi}_n$ is asymptotically zero-mean normal and its asymptotic covariance matrix is :

$$(I^{\alpha\alpha})^{-1}\,[I^{\alpha\alpha} - (I^{\alpha\alpha} \vdots I^{\alpha\beta})\, J (I^{\alpha\alpha} \vdots I^{\alpha\beta})' + J^{\alpha\alpha}]\,(I^{\alpha\alpha})^{-1}$$

$$= (I^{\alpha\alpha})^{-1} + (I^{\alpha\alpha})^{-1} J^{\alpha\alpha} (I^{\alpha\alpha})^{-1} - [I_{K_1} \vdots (I^{\alpha\alpha})^{-1} I^{\alpha\beta}]\, J [I_{K_1} \vdots (I^{\alpha\alpha})^{-1} I^{\alpha\beta}]'$$

Using the equality $(I^{\alpha\alpha})^{-1} I^{\alpha\beta} = - I_{\alpha\beta} I_{\beta\beta}^{-1}$, we get :

$$= (I^{\alpha\alpha})^{-1} + (I^{\alpha\alpha})^{-1} J^{\alpha\alpha} (I^{\alpha\alpha})^{-1} - J_{\alpha\alpha} + I_{\alpha\beta} I_{\beta\beta}^{-1} J_{\beta\alpha} + J_{\alpha\beta} I_{\beta\beta}^{-1} I_{\beta\alpha}$$
$$- I_{\alpha\beta} I_{\beta\beta}^{-1} J_{\beta\beta} I_{\beta\beta}^{-1} I_{\beta\alpha}$$

and, from the expression of $V_{as}\ \xi_n$ given in appendix 3 ,

$$= (I^{\alpha\alpha})^{-1} + (I^{\alpha\alpha})^{-1} J^{\alpha\alpha} (I^{\alpha\alpha})^{-1} - J_{\alpha\alpha} + V_{as}\ \xi_n - I_{\alpha\alpha} + I_{\alpha\beta} I_{\beta\beta}^{-1} I_{\beta\alpha}$$
$$+ J_{\alpha\beta} J_{\beta\beta}^{-1} J_{\beta\alpha}$$
$$= V_{as}\ \xi_n + (I^{\alpha\alpha})^{-1} J^{\alpha\alpha} (I^{\alpha\alpha})^{-1} - (J^{\alpha\alpha})^{-1}$$

REFERENCES

BOURGUIGNON, F. and B., SABATIER-LABEYRIE (1983)
Le programme QUALIT, DP 62, Ecole Normale Supérieure.

CHESHER, A.D. and M., IRISH (1984)
Residuals and diagnostics for Probit, Tobit and related models, University of Bristol, Discussion Paper 84-152.

COX, D.R. and E.J. SNELL (1968)
A general definition of residuals, JRSS B30, 248-275.

FOUTZ, R.V. and R.C. SRIVASTAVA (1977)
The performance of the likelihood ratio test when the model is incorrect, Annals of Statistics, 5, 1183-1194.

GOURIEROUX, C., MONFORT, A., RENAULT, E. and A. TROGNON (1984a)
Résidus généralisés ou interprétations linéaires de l'économétrie non linéaire, Document de travail ENSAE-INSEE n°8410.

GOURIEROUX C., MONFORT A., and TROGNON A. (1984b)
Pseudo likelihood methods : a survey, Document de travail ENSAE-INSEE n° 8406.

JENNRICH, R.I. (1969)
Asymptotic Properties of Non Linear Least Squares Estimations, Annals of Mathematical Statistics, 40 n°2, 633-643.

LANCASTER, T. and A.D. CHESHER (1984)
Residual analysis for censored duration data, University of Bristol, Discussion Paper 84-159.

MONFORT, A. (1982)
Cours de statistique mathématique, Economica, Paris.

TROGNON A. (1983)
Pseudo Asymptotic tests based on linear exponential family, Document de travail ENSAE-INSEE n° 8305.