ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 6/7 — 1987

Dynamic Error-in-Variables Models and Limited Information Analysis

Jean-Pierre FLORENS,
Michel MOUCHART, Jean-François RICHARD *

ABSTRACT. – A vector stochastic process x_t may be decomposed in to its expectation ξ_t and a residual process v_t. A linear dynamic model is defined by a set of dynamic linear relations constraining the ξ_t's given some conditioning variables and by the distribution of the v_t process. This paper presents a strategy for the specification of this class of models providing computable posterior distributions for a suitable class of prior measures. Some conditional independence properties characterizing exogeneity conditions through global or sequential cuts, innovation property or non causality relations are studied and are shown to allow reductions by conditioning of the model.

Modèles dynamiques à erreurs sur les variables et analyse en information limitée

RÉSUMÉ. – Un processus stochastique vectoriel x_t peut être décomposé en son espérance ξ_t et un résidu v_t. Un modèle linéaire dynamique est défini par un ensemble de relations dynamiques contraignant les ξ_t's pour des variables conditionnantes données et par la loi du processus v_t. Cet article présente une stratégie pour la spécification de cette classe de modèles fournissant les lois a posteriori calculables pour une classe de lois *a priori* convenable. On montre que le modèle peut être réduit en utilisant quelques propriétés d'indépendance conditionnelle caractérisant les conditions d'exogénéité par des coupes globales ou séquentielles, la propriété d'innovation ou des relations de non causalité.

* J. P. FLORENS : GREMAQ et Laboratoire de Statistique et de Probabilité. Université de Toulouse-I et III, place Anatole-France, 31042 Toulouse Cedex, France; M. MOUCHART : CORE, Université Catholique de Louvain, 34, voie du Roman-Pays, 1348 Louvain-la-Neuve, Belgique; J.-F. RICHARD : Duke University, NC, USA.

1 Introduction

In this paper we have three main objectives. First of all, we want to present a class of dynamic specifications suitable for modelling economic time series in the spirit of a limited information approach. Secondly we analyze some conditional independence properties which provide a coherent tool for general definitions of different kinds of exogeneity, noncausality, innovation properties... These definitions are applied to the class of dynamic linear models we have defined and are translated, for this class of models, into parametric restrictions. At last we propose a bayesian analysis of the structural linear dynamic model and a class of prior probabilities is given which keeps the numerical integration problem within a reasonable size.

In this introduction a general presentation of the sampling process of the structural linear dynamic model is given but we concentrate our attention in Section 2 to a particular case : the specification of the prior and the analysis of the posterior are only made in the example. The extension to a more general case only requires further technicalities; the discussion in a simple case is more illuminating. The section 3 deals with the general model and exposes abstract definition and specific characterizations of exogeneity, non causality... viewed as conditions allowing some reductions of the model without losing any relevant information.

Let us consider two vectors of observations, $x=(x_1', \ldots, x_T')'$ where any x_t is an element of $\mathbb{R}^m$ and $w=(w_{1-s}', \ldots, w_0', w_1, \ldots, w_T)'$ which is a sequence of realizations of an unspecified conditioning stochastic process. The specified model is the model generating x given w and given a vector of parameters denoted λ. The assumptions are the following:

(1) $$x \mid w, \lambda \sim \mathrm{N}(\tilde{\xi}, \tilde{\mathrm{V}})$$

where

$$\tilde{\xi}=\mathrm{E}(x \mid w, \lambda)=(\xi_1', \ldots, \xi_T')', \qquad \xi_t \in \mathbb{R}^m$$

and

$$\tilde{\mathrm{V}}=\mathrm{V}(x \mid w, \lambda) \text{ is an } m\mathrm{T} \times m\mathrm{T} \text{ PDS matrix}$$

$\exists\, \mathrm{B}_\theta(\mathrm{L})$ and $\mathrm{C}_\theta(\mathrm{L})$ two matrices of polynomials in the lag operator L such that:

(2) $$\forall t, \quad \mathrm{B}_\theta'(\mathrm{L})\,\xi_t + \mathrm{C}_\theta'(\mathrm{L})\,w_t = 0$$

$$\mathrm{B}_\theta(\mathrm{L})\; m \times p \qquad \text{and} \qquad \mathrm{C}_\theta(\mathrm{L})\; l \times p$$

$$\text{rank of } \mathrm{B}_\theta^0 = p$$

B_θ and C_θ depend on a parameter θ and λ is here the union of θ, the unknown elements of $\tilde{\mathrm{V}}$ and the ξ_t verifying the constraint (2). B_θ^0 is the matrix of the constant terms of $\mathrm{B}_\theta(\mathrm{L})$. Moreover we assume that the orders of the polynomials in $\mathrm{B}_\theta(\mathrm{L})$ and $\mathrm{C}_\theta(\mathrm{L})$ are known and that the values of $\xi_0, \xi_{-1}, \ldots$ necessary in (2) are given. By convention we incorporate theses values in the $w_0, w_{-1}, \ldots$ Hence if $w_t \in \mathbb{R}^l$ if $t \geqslant 1$, $w_t \in \mathbb{R}^{l+m}$ for $t<1$. s is determined by the maximal order of the lag polynomials. We have assumed the regularity of $\tilde{V}$ but this assumption does not restrict the model because if, for example, a variable in x_t is not random (in which case if a column and a row of $\tilde{V}$ are equal to zero) this variable will be included w_t. More generaly any singularity in $\tilde{V}$ would be eliminated by reducing the size of x_t and increasing the size of w_t.

In the last section we will use a more compact notation of the model: the set of relations $B'_\theta(L)\,\xi_t + C'_\theta(L)\,w_t = 0$, $t=1, \ldots, T$ can be written:

$$\tilde{B}'_\theta \xi + \tilde{C}'_\theta w = 0 \tag{3}$$

where $\tilde{B}_\theta$ is a $mT \times pT$ matrix and $\tilde{C}$ is a $[lT+(l+m)(s+1)] \times pT$ matrix which are easily constructed given $B_\theta(L)$ and $C_\theta(L)$. These definitions will be illustrated later on (see in particular the example 1).

An interpretation of this kind of models could be the following. The sequence of observations x_t is decomposed in the sum $\xi_t + v_t$ where ξ_t is the (conditional) expectation of x_t and v_t is defined by $x_t - \xi_t$ and then has a zero expectation. Typically it is assumed that v_t is a stationary process and that ξ_t is non-stationary but this interpretation is not necessary. The hypothesis about the v_t process are used to provide $\tilde{V}$ with some structure in order to allow a parsimonious parametrization of $\tilde{V}$. For instance if v_t is an iid process, $\tilde{V} = I_T \otimes V$ where V is the variance of v_t.

In the dynamic case, it will appear that the distributional assumptions which simplify the treatment of the model are not directly expressed in terms of v_t but in terms of a transformation of v_t by a linear filter. This transformation will be assumed stationary with a particular lag structure. This point will be developped in Section 2.

The ξ_t are the sequence of the expectations of x_t given the parameters and w and are constrained by the equations (2). Still we do not assume at this stage that B_θ is square. Then ξ_t cannot be expressed as a function of w_t but have $m-p=q$ free elements for any t. Equations (2) may be considered as the "structural" equations of the economic theory which is described by the statistical model and the unconstrained components of the ξ_t are called "incidental" and new incidental values appear at each observation.

In the theory of the error $-$in$-$ variables models such models are called "structural" if the incidental values are considered as (unobservable) realizations of a stochastic process the parameters of which are included in the parameter of the models. These models are called "functional" if the incidental values are treated as parameters. The classical treatment$-$for example the maximum likelihood estimation$-$of these two kinds of models is different: usually the first one has a fixed finite dimension parameter set but the second one has a parameter space whose dimension increases with the sample size. Let us remark however that the asymptotic analysis of the functional model requires, at least implicitly, some assumptions about the process generating the incidental parameters. In our presentation of the linear model, we will treat the incidental values as parameters but in our bayesian approach we provide this sequence of parameters with a prior distribution. Then the difference between the two approaches reduces to a divergence between two interpretations of this prior which can be interpreted either as a sampling process on unobservables values or as a probabilistic translation of subjective prior knowledges.

The generality of the statistical structure we have defined is one of its interest. The two main restrictions of our specification are the normality of the residuals and the linearity of the relations and it is well-known that both may be viewed as approximations of a more complex reality. Despite these limitations the model covers a wide range of situations such as (dynamic) simultaneous equations, (dynamic) error-in-invariables models and some (dynamic) factor analysis models.

A complete bibliography on these topics is not be given in this paper (see, for more references the survey of GRILICHES [1974] or the recent paper of DEISTLER [1986]).

2 Analysis of a One Equation Model

2.2. Specification of the Model

We shall concentrate our attention on a particular case essentially characterized by the number of variables which is equal to two and the number of structural equations which is equal to one. We denote $x_t = (x_{t1}, x_{t2})'$ the vector of explained variables and the vector of the conditioning variables is an element of $\mathbb{R}^l$ for $t \geqslant 1$. The conditional expectations $\xi_t = (\xi_{t1}, \xi_{t2})'$ are related by the dynamic relationship:

$$\beta_1(L)\,\xi_{t1} + \beta_2(L)\,\xi_{t2} + \gamma(L)'\,w_t^1 = 0 \tag{4}$$

where $\beta_1(L)$, $\beta_2(L)$ and $\gamma(L)$ are and a polynomials two vector of polynomials in the lag operator L with known orders and $\beta_1(L)$ is normalized by its constant term assumed to be equal to one. w_t' is a sub vector of w_t. We shall denote by θ the vector of the unknown parameters of β_1, β_2 and γ.

An "instrumental" relation will also be introduced in the following way. We assume that a known dynamic linear relation of the ξ_{t1} and ξ_{t2} is a linear function of another subvector w_t^2 of w_t. More precisely we introduce a relation $\eta_t + \delta'(L)\,w_t^2 = 0$ where $\delta(L)$ is a vector of polynomials in the lag operator and η_t is defined by $b_1^\theta(L)\,\xi_{t1} + b_2^\theta(L)\,\xi_{t2}$; b_1^θ and b_2^θ are known functions of θ. For instance one can choose $b_1(L) = 0$ and $b_2(L) = 1$ but our presentation is not restricted to this particular case.

Let us note that, at this stage of modelling, the introduction of this second relation is not a supplementary assumption if no condition about the ranq of the matrix $(w_1^2, \ldots, w_T^2)$ is required. One can define: $b_1^\theta(L) = 0$, $b_2^\theta(L) = 1$, $w_t^2 \in \mathbb{R}^T$ to be the t-th column of the identity matrix I_T

and $\delta'(L)=(\delta_1, \ldots, \delta_T)$, polynomials of order 0. Then this second relation reduces to $\xi_{t2}=\delta_t$ and is simply a redefinition of ξ_{t2}.

The model can be written:

$$\begin{pmatrix} x_{t1} \\ x_{t2} \end{pmatrix} = \begin{pmatrix} \xi_{t1} \\ \xi_{t2} \end{pmatrix} + \begin{pmatrix} v_{t1} \\ v_{t2} \end{pmatrix} \tag{5}$$

$$\begin{pmatrix} \beta_1(L) & \beta_2(L) \\ b_1^\theta(L) & b_2^\theta(L) \end{pmatrix} \begin{pmatrix} \xi_{t1} \\ \xi_{t2} \end{pmatrix} + \begin{pmatrix} \gamma'(L) & 0 \\ 0 & \delta'(L) \end{pmatrix} w_t = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{6}$$

λ is the set of the parameters involving θ, the parameters of δ and of the distribution of v_t.

Let us remark that w_t is the union of w_t^1 and w_t^2 but these two vectors may have common elements. In some cases, $w_t^2=w_t$ and w_t^1 is a sub vector of w_t^2.

We do not introduce now the hypothesis about the distribution of v_t; these hypothesis will be more easily expressed after a transformation of the variables.

For a given value of θ, let us define the following transformation:

$$\forall t=1, \ldots, T,$$

$$\begin{pmatrix} x_{t1} \\ x_{t2} \end{pmatrix} \to \begin{pmatrix} u_t^\theta \\ \psi_t^\theta \end{pmatrix} = \begin{pmatrix} \beta_1(L)\,x_{t1} + \beta_2(L)\,x_{t2} + \gamma'(L)\,w_t^1 \\ b_1^\theta(L)\,x_{t1} + b_2^\theta(L)\,x_{t2} \end{pmatrix} \tag{7}$$

Given θ this transformation defines an affine transformation of $(x_1', x_2')' \in \mathbb{R}^{2T}$ into $(u^{\theta\prime}, \psi^{\theta\prime})' \in \mathbb{R}^{2T}$ which can be written as:

$$\begin{pmatrix} u^\theta \\ \psi^\theta \end{pmatrix} = \tilde{Q}^\theta \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \tilde{q}^\theta \tag{8}$$

where $\tilde{Q}^\theta$ is $2T \times 2T$ regular and $\tilde{q}^\theta \in \mathbb{R}^{2T}$.

If $l(x \mid w, \lambda)$ and $l^*(u^\theta, \psi^\theta \mid w, \lambda)$ are respectively the sampling densities of x and (u^θ, ψ^θ), we have the following relation:

$$l(x \mid w, \lambda) = \| \tilde{Q}^\theta \| \, l^*(u^\theta, \psi^\theta \mid w, \lambda) \tag{9}$$

and the distributional assumptions can be made equivalently on x or on (u^θ, ψ^θ).

More precisely, this distribution will be specified in two steps. We first consider the distriburion of u^θ and in the second stage the distribution of ψ^θ given u^θ.

(i) Conditionally on (w, λ) the vector of structural residuals u_t^θ is assumed to be normal with a zero mean and a variance $\tilde{\Sigma}_\rho$ where ρ is the vector of the parameter of the u_t^θ-process. As we shall treat numerically ρ, the specification of $\tilde{\Sigma}_\rho$ is very flexible under the condition of keeping the dimension of the structural parameters (θ, ρ) at a size making feasible a numerical maximisation or a numerical integration. Typicallyt u_t^θ will be modelled as an ARMA process with low degree polynomials.

(ii) The model generating ψ^θ given u^θ will be specified in a more restrictive wag.

This model can be characterized by the distribution of ψ_t^θ given its past and given the whole trajectory of the (u_t^θ). Must also be introduced as conditioning variables the w_t^2. This assumption can be described by the following relation:

$$\psi_t^\theta = \mu(L)\,\psi_t^\theta + \lambda(L)\,u_t^\theta + \delta'(L)\,w_t^2 + \varepsilon_t \tag{10}$$

where

$$\mu(L) = \sum_{i=1}^{r} \mu_i L^i,$$

$\delta(L)$: have been introduced above and

$$\lambda(L) = \sum_{i=-k_1}^{k_2} \lambda_i L^i \qquad (r,\ k_1,\ k_2 \in \mathbb{N},\ \text{known degrees}).$$

ε_t is a normal white noise of variance equal to ω independent by definition from its past and from the sequence of $(u_t^\theta)_t$. Let us note that (10) involves past and future values of u_t^θ but the dependence of ψ_t from u_t^θ is restricted to a finite past and a finite future. On the same line, ψ_t^θ depends on its own past only through a known finite number of lags.

Let us define $\bar{\mu}' = (\mu_1, \ldots, \mu_r)'$, $\bar{\lambda}' = (\lambda_{-k_2}, \ldots, \lambda_{k_1})'$, $\bar{\delta}$ the vector of all the coefficients of the polynomials in $\delta(L)$. By means of λ a suitable definition of $\bar{\psi}^\theta$, $\bar{U}^\theta$ and $\bar{W}^2$, equations (10) may be expressed by the following matrix form:

$$\psi^\theta = \bar{\psi}^\theta \bar{\mu} + \bar{U}^\theta \lambda + \bar{W}^2 \delta + \varepsilon. \tag{11}$$

For instance a row of $\bar{U}^\theta$ has the form $(u_{t+k}^\theta, \ldots, u_{t-k}^\theta)$. More over:

$$\varepsilon \mid u^\theta,\ w,\ \lambda \sim N(0,\ \omega I_T) \tag{12}$$

We may now give a precise description of the complete the parameter:

$$\lambda = (\theta,\ \rho,\ \bar{\mu},\ \bar{\lambda},\ \bar{\delta},\ \omega)$$

and we further assume that $\lambda_1 = (\theta, \rho)$ and $\lambda_2 = (\bar{\mu}, \bar{\lambda}, \bar{\delta}, \bar{\omega})$ are variation free. In the bayesian analysis the prior probability on λ_1 and λ_2 will make these two vectors independent. Moreover we will assume that λ_2 is un constrained and the conditional model (11) will may be treated analogously to a regression model.

The likelihood can be written in the following way:

$$\begin{aligned} l(y \mid w, \lambda) &= \|\tilde{Q}^\theta\|\, l^*(u^\theta \mid w, \lambda_1)\, l^*(\psi^\theta \mid u^\theta, w, \lambda_2) \\ &= \|\tilde{Q}^\theta\|\,(2\pi)^{-T/2}\,|\tilde{\Sigma}_\rho|^{-1/2} \exp -\frac{1}{2} u^{\theta\prime}\,\tilde{\Sigma}_\rho^{-1}\,\tilde{u}^\theta \\ &\quad \times (2\pi)^{-T/2}\,\omega^{-T} \exp -\frac{1}{2\omega}(\psi^\theta - \bar{\Psi}^\theta \bar{\mu} - \bar{U}^\theta \bar{\lambda} - \bar{W}^2 \bar{\delta})' \\ &\quad \times (\psi^\theta - \bar{\psi}^\theta \bar{\mu} - \bar{U}^\theta \bar{\lambda} - \bar{W}^2 \bar{\delta}). \end{aligned} \tag{13}$$

We have followed the general philosophy of the limited information approach: a maximum flexibility is preserved for the specification of the structural part of the model (the equation (4) and the residuals u_t^θ) but the "instrumental" equation (10) is restricted to have a finite number of lagged and future values, white noise residuals and parameters which are unrelated to the parameters of the structural part of the specification.

This presentation follows a line introduced by the authors in the static model in FLORENS *et al.* [1976 and 1979]. Some dynamic extensions was realized by RICHARD [1984] and FLORENS *et al.* [1986].

A alternative approach was presented in the static model by HOLLY and SARGAN [1982] and HOLLY [1983] and these authors use a decomposition of the sampling process in the marginal process generating ψ_t^θ and the conditional process of u_t^θ given ψ^θ (in our notation). In the static model the different approaches are equivalent but in the dynamic case it should be noted that, for instance, a finite lag structure on the model generating ψ_t^θ given u^θ is not equivalent to a finite lag structure on the v_t and is not equivalent to a finite lag structure on the model generating u_t^θ given ψ^θ. Moreover the construction of the distribution of v_t given the distribution of u_t^θ and of ψ_t^θ given u^θ and the decomposition into a marginal process of ψ_t^θ and a conditional process of u_t^θ given ψ^θ are very tedious in the dynamic model.

2.2. Inference

Although we mainly develop the bayesian inference we first give briefly the main steps for the computation of the maximum likelihood estimator. We first assume $\bar{W}^2$ of fixed rank. This computation is done in two steps. In the first step, the maximisation is performed assuming (θ, ρ) given and these reduce to the computation of the maximisation of the likelihood of the linear regression model (11). We obtain:

$$(14) \qquad \begin{pmatrix} \hat{\bar{\mu}}(\theta) \\ \hat{\bar{\lambda}}(\theta) \\ \hat{\bar{\delta}}(\theta) \end{pmatrix} = \begin{pmatrix} \bar{\Psi}^{\theta\prime}\bar{\Psi}^\theta & \bar{\Psi}^{\theta\prime}\bar{U}^\theta & \bar{\Psi}^{\theta\prime}\bar{W}^2 \\ \bar{U}^{\theta\prime}\bar{\Psi}^\theta & \bar{U}^{\theta\prime}\bar{U}^\theta & \bar{U}^{\theta\prime}\bar{W}^2 \\ \bar{W}^{2\prime}\bar{\Psi}^\theta & \bar{W}'\bar{U}^\theta & \bar{W}^{2\prime}\bar{W}^2 \end{pmatrix}^{-1} \begin{pmatrix} \bar{\Psi}^{\theta\prime}\psi^\theta \\ \bar{U}^{\theta\prime}\psi^\theta \\ \bar{W}^{2\prime}\psi^\theta \end{pmatrix}$$

and

$$(15) \qquad \hat{\omega}(\theta) = \frac{1}{T}\psi^{\theta\prime} M_{\bar{\Psi}^\theta\bar{U}^\theta\bar{W}^2}\psi^\theta$$

with $M_Z = I - Z(Z'Z)^{-1}Z'$.

In the second step we compute the concentrated likelihood:

$$(16) \qquad \mathrm{Log}\, l(y \mid w, \theta, \rho) = \mathrm{Log}\,\|\tilde{Q}^\theta\| - \frac{1}{2}\mathrm{Log}\,|\tilde{\Sigma}_\rho| - \frac{1}{2}u'^\theta\tilde{\Sigma}_\rho u^\theta$$

$$- \frac{T}{2}\mathrm{Log}\,\psi^{\theta\prime} M_{\bar{\Psi}^\theta\bar{U}^\theta\bar{W}^2}\psi^\theta + \text{constant}.$$

A numerical maximisation of (16) gives the estimators (θ, ρ).

In some particular cases, more can be done analytically. For example, let us assume that $\tilde{Q}^\theta$ and ψ^θ only depend on the parameters of $\beta_1(L)$ and $\beta_2(L)$ and that the parameters of $\gamma(L)$ are not constrained. Then these last parameters may be estimated given $\beta_1(L)$ and $\beta_2(L)$ by least squares estimators using $-\beta_1(L)x_{t1}-\beta_2(L)x_{t2}$ as the dependant variable and w_t^1 and its lagged values as the independant variables. The concentrated likelihood is then computed without any difficulties. The fact that $\bar{W}^2$ has a fixed rank restricts this presentation to a limited information framework and the classical analysis of a dynamic error-in-variables model is begand the scope of this paper. However the bayesian approach—at least in the case of proper prior distributions—does not require any assumption on the rank of $\bar{W}^2$.

We now consider the bayesian inference and first specify the prior distribution we shall use. This prior measure is defined in two parts. As θ and ρ will be treated numerically, the prior probability on (θ, ρ) will not be explicitely given. Its density with respect to Lebesgue measure is denoted by $m(\theta, \rho)$. The prior specification on the parameters $(\bar{\mu}, \bar{\lambda}, \bar{\delta}, \omega)$ is more restrictive and, in order to integrate out these parameters in the posterior density, the most convenient choice is a natural conjugate prior probability associated to the conditional model (11) which has the same structure as a linear regression model. More formally $(\bar{\mu}, \bar{\lambda}, \bar{\delta}, \omega)$ is provided with an inverted gamma normal distribution. i.e.:

(17) $$\omega \mid w, \theta, \rho \sim I\Gamma(\omega_0, \nu_0)$$

(18) $$\begin{pmatrix} \bar{\mu} \\ \bar{\lambda} \\ \bar{\delta} \end{pmatrix} \Bigg| w, \theta, \rho, \omega \sim N\left[\begin{pmatrix} \bar{\mu}_0 \\ \bar{\lambda}_0 \\ \bar{\delta}_0 \end{pmatrix}, \omega L_0^{-1}\right)$$

In general the hyper parameter of this prior can be functions of (θ, ρ) and of the conditioning variables w. The posterior probability on $(\bar{\mu}, \bar{\lambda}, \bar{\delta}, \omega)$ has the same distribution; its the hyperparameters are equal to:

(19) $$\nu_* = \nu_0 + T$$

(20) $$L_* = L_0 + (\bar{\psi}^\theta\ \bar{U}^\theta\ \bar{W}^2)'(\psi^{-\theta}\ \bar{U}^\theta\ \bar{W}^2)$$

(21) $$\begin{pmatrix} \bar{\mu}_* \\ \bar{\lambda}_* \\ \bar{\delta}_* \end{pmatrix} = L_*^{-1}\left[L_0\begin{pmatrix} \bar{\mu}_0 \\ \bar{\lambda}_0 \\ \bar{\delta}_0 \end{pmatrix} + (\bar{\psi}^\theta\ \bar{U}^\theta\ \bar{W}^2)'\psi^\theta\right]$$

(22) $$\omega_* = \omega_0 + (\psi^\theta - \bar{\psi}^\theta \bar{\mu} - \bar{U}^\theta \bar{\lambda}_0 - \bar{W}^2 \delta_0)'[I_T - (\bar{\psi}^\theta\ \bar{U}^\theta\ \bar{W}^2) L_*^{-1} (\bar{\psi}^\theta\ \bar{U}^\theta\ \bar{W}^2)'] (\psi^\theta - \bar{\psi}^\theta \bar{\mu}_0 - \bar{U}^\theta \lambda_0 - \bar{W}^2 \delta_0)$$

The predictive density of ψ^θ given $\bar{\psi}^\theta$, $\bar{U}^\theta$, $\bar{W}^2$ and (θ, ρ) is given by:

(23) $$p(\psi^\theta \mid \bar{\psi}^\theta, \bar{U}^\theta, \bar{W}, \theta, \varphi) \propto \omega_0^{(1/2)\nu_0} \mid L_0 \mid^{1/2} \omega_*^{-(1/2)\nu_*} \mid L_* \mid^{1/2}$$

Details on these computations can be found e.g. in De Groot [1970] and are recalled in Florens Mouchart [1986].

The posterior probability on (θ, ρ) is then equal to:

$$(24) \qquad m(\theta, e \mid x, w) \propto m(\theta, e) \| \tilde{Q}^{\theta} \| \, |\Sigma_e|^{-1/2} \exp -\frac{1}{2} u^{\theta\prime} \tilde{\Sigma}_e^{-1} u^{\theta}$$

$$\times \omega_0^{(1/2)\nu_0} |L_0|^{1/2} \omega^{-(1/2)\nu_*} |L_*|^{-1/2}$$

As in classical analysis, the treatment of this posterior density must be performed numerically but further analytical integrations may be realised at the coast of introducing additional assumptions.

For instance if u_t^{θ} is assumed to be a normal white noise with variance equal to α (here the parameters ρ reduce to α) a suitable prior on α is an inverted gamma distribution $I\Gamma(\alpha_0, k_0)$ given θ(α_0 may be a function of θ). In this case the posterior density becomes:

$$(25) \quad m(\theta \mid x, w) \propto m(\theta) \| Q_{\theta} \| \, \omega_0^{(1/2)\nu_0} |L_0|^{1/2} \omega_*^{-(1/2)\nu_*}$$

$$\times |L_*|^{-1/2} \alpha_0^{(1/2)\nu_0} \alpha_*^{-(1/2)k_*}$$

where $\alpha_* = \alpha_0 + u'^{\theta} u^{\theta}$ and $k_* = k_0 + T$.

Remark: In the posterior distribution (24) [and in the special case (25)] ω_* and L_* are complicated functions of θ. These expressions may be simplified using some matrix manipulations. We just consider a particular case in which ψ_t^{θ} does not depend on θ and is then a known function of the observable variables. We can rewrite (11) as:

$$(26) \qquad \psi^{\theta} = \bar{U}^{\theta} \bar{\lambda} + \bar{Z} \bar{\pi} + \varepsilon$$

where $\bar{Z} = (\bar{\psi}^{\theta}, \bar{W}^2)$ and $\bar{\pi} = (\bar{\mu}\, \delta)$ and all formulae are rearranged according to (26). L_0 can now be written as:

$$(27) \qquad L_0 = \begin{pmatrix} H_0 + E_0' G_0 E_0 & E_0' G_0 \\ G_0 E_0 & G_0 \end{pmatrix}.$$

where the dimension of the square matrix H_0 is the dimension of $\bar{\lambda}$(i.e. $k_1 + k_2 + 1$), and H_0, G_0 and E_0 are unambigously characterized by (27).

We now define:

$$(28) \qquad \Phi_0 = \begin{pmatrix} H_0 & H_0 \bar{\lambda}_0 \\ \bar{\lambda}_0' H_0 & \omega_0 + \bar{\lambda}_0' H_0 \bar{\lambda}_0 \end{pmatrix}$$

$$(29) \qquad P_0 = (E_0\, \bar{\pi}_0'\, E_0 + \bar{\pi}_0), \qquad \bar{\pi}_0 = \begin{pmatrix} \bar{\mu}_0 \\ \bar{\delta}_0 \end{pmatrix}$$

$$(30) \qquad \Phi_* = \Phi_0 + [(\bar{U}^{\theta}\ \psi^{\theta}) - \bar{Z} P_0]' (I_T - \bar{Z} G_*^{-1} \bar{Z}') [\bar{U}^{\theta}\ \psi^{\theta}) - \bar{Z} P_0]$$

$$(31) \qquad H_0 = H_0 + (U^{\theta} - \bar{Z} E_0)' (I_T - \bar{Z} G_*^{-1} \bar{Z}') (U^{\theta} - \bar{Z} E_0)$$

$$(32) \qquad G_* = G_0 + \bar{Z}' \bar{Z}.$$

Then (see FLORENS *et al.* [1986] appendix B)

$$(33)\quad p(\psi^\theta \mid \bar{\psi}, \bar{u}^\theta, \bar{w}, \theta, \rho) \propto |\Phi_0|^{(1/2)\nu_0} |G_0|^{1/2} \times |H_0|^{-(1/2)(\nu_0-1)} |\Phi_*|^{-(1/2)\nu_*} |G_*|^{-1/2} |H_*|^{(1/2)(\nu_*-1)}.$$

We conclude this section by giving the posterior distribution of a very simple example. Let us consider a bivariate stochastic process x_t such that $\xi_t = E(x_t)$ satisfies the relations:

$$(34)\qquad \xi_{t\,1} + \beta_1 \xi_{t-1,1} + \beta_2 \xi_{t\,2} = 0$$

$$(35)\qquad \xi_{t\,2} + \delta w_t = 0$$

where w_t is an observable conditionning process. We assume moreover that

$$u_t^\theta = x_{t\,1} + \beta_1 x_{t-1,2} + \beta_2 x_{t,2}$$

is a white noise and in this example $\psi_{t,2}^\theta = x_{t,2}$. The conditional distribution of $x_{t,2}$ given its past and the sequence of u_t^θ is furthermore assumed to be characterized by:

$$(36)\qquad x_{t\,2} = \mu x_{t-1\,2} + \lambda_1 u_t^\theta + \lambda_2 u_{t-1}^\theta + \delta w_t + \varepsilon_t$$

where ε_t is also by construction a white noise. The variances of u_t^θ and ε_t are respectively α and ω.

Equation (36) may be written as in (26), i. e.:

$$(37)\qquad x_2 = \bar{V}^\theta \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} + \bar{Z} \begin{pmatrix} \mu \\ \delta \end{pmatrix} + \varepsilon$$

where

$$\bar{U}^\theta = \begin{pmatrix} u_1^\theta & \dots & u_T^\theta \\ u_0^\theta & \dots & u_{T-1}^\theta \end{pmatrix}' \quad \text{and} \quad \bar{Z} = \begin{pmatrix} x_{02} & \dots & x_{T-1\,2} \\ w_1 & \dots & w_T \end{pmatrix}'$$

Let us specify a "non informative" prior measure on $(\lambda_1, \lambda_2, \mu, \delta, \omega)$ characterized by a density equal to $\omega^{-(1/2)(\nu_0+6)}$ (the general form of the exponant is an arbitrary number plus the number of regression parameters (here equal to 4) plus 2).

The density of x_2 given $\bar{U}^\theta$, $\bar{Z}$ and $(\beta_1, \beta_2, \alpha)$ is then computed:

$$(38)\quad p(x_2 \mid \bar{U}^\theta, \bar{Z}, \beta_1, \beta_2, \alpha) = p(x_2 \mid \bar{U}^\theta, \bar{Z}, \beta_1, \beta_2)$$
$$\propto (x_2' M_{\bar{U}^\theta \bar{Z}} x_2)^{-(1/2)(T+\nu_0)} |(\bar{U}^\theta\ \bar{Z})'(\bar{U}^\theta\ \bar{Z})|^{-1/2}$$
$$\propto |(\bar{U}^\theta\ x_2)' M_{\bar{Z}} (\bar{U}^\theta\ x_2)|^{-(1/2)(T+\nu_0)} |\bar{U}^{\theta\prime} M_{\bar{Z}} \bar{U}^\theta|^{(1/2)(T+\nu_0-1)}$$

In order to compute the posterior density of (β_1, β_2) we assume that α_0 does not depend on (β_1, β_2) and we remark that $\|Q^\theta\| = 1$. Then we get:

$$(39)\quad m(\beta_1, \beta_2 \mid x, w) \propto m(\beta_1, \beta_2)(\alpha_0 + u^{\theta\prime} u^\theta)^{-(1/2)(T+\mu_0)} p(x_2 \mid \bar{U}^\theta, \bar{Z}, \beta_1, \beta_2)$$

where $u^\theta = (u_1^\theta, \dots, u_T^\theta)'$.

3. Conditional Independences in Dynamic Linear Model

In several recent papers, precise definitions of noncausality and of exogeneity have been proposed together with an analysis of the connections between these concepts (see e. g., FLORENS *et al.* [1980], ENGLE *et al.* [1983] and FLORENS MOUCHART [1985, 1986] in which more references are given). We will apply these definitions in the dynamic linear model studied here, and also discuss an innovation property which is standard in time series analysis.

Each property will be expressed by a conditional independence property which can be considered at two different stages of the model specification. First, such an independence may be assumed when the model is build and implies a simplification of the statistical model (for instance an exogenity assumption will imply that the model of the process generating the exogenous variables does not need to be specified and the dimension of the parameter space is then reduced). A different use of these independance conditions is obtained if they are not assumed at the beginning of the model building but if they are tested through a restriction on the parameter space equivalent to the conditional independence of interet.

We will consider four different concepts: the global cut and the sequential cut which both formalize the intuitive concept of exogeneity, the noncausality property and the innovation property. For each of these concept we will first recall the general definition and then apply it to the linear model.

3.1. Global Cut

Let x be the vector of the observations of a model. The likelihood function $l(x \mid \lambda, \omega)$ depends on parameters λ and on observed conditioning variables w. Let x be partionned into (y, z). This partition realizes *a global cut* of the model (given w) if there exists a partition of λ in (λ_c, λ_m) such that:

(i) λ_c and λ_m are sufficient parameters for the model generating y given z and for the model generating z respectively. More specifically:

(40) $$l(y \mid z, w, \lambda) = l(y \mid z, w, \lambda_c)$$

(41) $$l(z \mid w, \lambda) = l(z \mid w, \lambda_m)$$

(ii) λ_c and λ_m are variation free.

If the parameter space is provided with a prior measure, the conditions (40) and (41) can be viewed as conditional independences: (40) means that y is independent of λ given λ_c, z and w and (41) means that z is independent of λ given λ_m and w.

The condition (ii) becomes an *a priori* independence in the bayesian approach.

Let us note that if this prior probability is formally necessary in order to use (conditional) independences, this conditions remain satisfied if the prior is replaced by another probability having the same null sets.

We will use the classical definition of the global cut essentially because we want in general to specify the prior after the reduction of the model on a parameter space as small as possible.

Let us recall the intuitive meaning of the global cut property.

Given w, the joint sampling distribution of (y, z) may always be decomposed into its marginal distribution of z and its conditional distribution of y given z but the independence between the parameters of these two probabilities (formalized through a variation free assumption or by a prior independence) characterizes the idea of a "structural" separation between the process generating z and the process generating y given z and gives a precise definition of "z is exogenous". The statistical implications of this property are obvious: λ_c and λ_m may be estimated without losing information using only the conditional distribution and the marginal distribution respectively. In particular the estimation of λ_c is robust to a modification of the marginal process provided that the cut property remains valid.

In order to apply the global cut concept to the dynamic linear model we will use the global notation of the model ((1) (3)) in which the linear constraint is written $\tilde{B}'_\theta \xi + \tilde{C}'_\theta w = 0$.

Partionning x into $(y', z')'$ and ξ in $(\xi'_y, \xi'_z)'$ this equation becomes $\tilde{B}^{y'}_\theta \xi_y + \tilde{B}^{z'}_\theta \xi_z + \tilde{C}'_\theta w = 0$ where $\tilde{B}'_\theta = (\tilde{B}^{y'}_\theta, B^{z'}_\theta)$. Let us recall that x has $\tilde{m} = m\mathrm{T}$ elements. If y has $\tilde{n}$ elements $\tilde{B}^y_\theta$ is a $\tilde{n} \times \tilde{p}$ matrix ($\tilde{p} = p\mathrm{T}$). The existence of n such as $\tilde{n} = n\mathrm{T}$ is not necessary and the number of exogenous variables may be not constant over the time. The global cut property is characterized by the following proposition.

PROPOSITION 1 : Let us assume Rank $\tilde{B}^y_\theta = \tilde{p}$. The partition $x' = (y', z')$ realizes a global cut of the model if and only if z and $u_\theta = \tilde{B}'_\theta x + C'_\theta w$ are independant given w and λ.

This proposition is prooved in the appendix. Under the assumption of independence between u_θ and z, the marginal model is

(42) $$z \mid \xi_z, \tilde{V}_{zz}, w \sim \mathrm{N}(\xi_z, \tilde{V}_{zz}), \qquad \xi_z \in \mathbb{R}^{\tilde{k}}, \quad \tilde{V}_{zz} \in \mathscr{C}_{\tilde{k}}$$

and the conditional model is

(43) $$y \mid \xi^z_y, \theta, \tilde{V}_{yy} w, z \sim \mathrm{N}(\xi^z_y, \tilde{V}_{yy\,.\,z})$$

where ξ^z_y satisfies:

(44) $$\tilde{B}'_\theta \begin{pmatrix} \xi^z_y \\ z \end{pmatrix} + \tilde{C}'_\theta w = 0$$

and

$$\tilde{V}_{yy\,.\,z} = \tilde{V}_{yy} - \tilde{V}_{yz} \tilde{V}^{-1}_{zz} \tilde{V}_{yz}.$$

The exogenity condition characterized by the global cut property is then used to substitute in the equation $B'_\theta \tilde{\xi} + C'_\theta w = 0$ the expectation of x conditional on z (and w and the parameters) for the expectation on x (given w and the parameters). The conditional model is formally identical to the original model but the z-variables are dropped in the w-variables without loss of information on the parameter θ of interest. Indeed condition (44) may be rewritten as:

$$\tilde{B}_\theta^{y'} \tilde{\xi}_y^z + (\tilde{B}_\theta^{z'}, \tilde{C}'_\theta) \begin{pmatrix} z \\ w \end{pmatrix} = 0. \tag{45}$$

In our presentation w is not necessarily assumed to be exogenous and some information may have been lost by conditioning on w. In a bayesian analysis the global cut will imply that z and the parameters of the conditional model are independent. Then if the prior is specified after the conditioning on z, it will not depend on z.

However the parameters of the prior distribution could depend on w and then take into account the information given by w on the parameters.

Remark: In order to simplify our presentation it has implicitely been assumed that $\tilde{V}_{zz}$ is not related to the parameters of the conditional model and that $\tilde{V}_{yy.z}$ is not related to the parameters of the marginal model.

Example 1: Let us assume $x_t \in \mathbb{R}^2$ $x_t = \begin{pmatrix} y_t \\ z_t \end{pmatrix}$ such that

$$\begin{cases} y_t = \xi_{yt} + v_{yt} \\ z_t = \xi_{zt} + v_{zt} \end{cases} \qquad t = 1, \ldots, T \tag{46}$$

where the expectations satisfy:

$$\xi_{yt} + \alpha \xi_{yt-1} + \beta \xi_{zt} + \gamma \xi_{zt-1} = 0. \tag{47}$$

The problem of specifying of the distribution of the initial conditions is treated by assuming the preceding relation true for $t = 2, \ldots, T$ and by considering the $\xi_{y\,1}$ and $\xi_{z\,1}$ as parameters.

Denoting $y = (y_1, \ldots, y_T)'$, $z = (z_1, \ldots, z_T)'$, $x' = (y', z')$, $\xi_y = (\xi_{y\,1}, \ldots, \xi_{y\,T})'$, $\xi_z = (\xi_{z\,1}, \ldots, \xi_{z\,T})'$, $\tilde{\xi}' = (\xi'_y, \xi'_z)$, the model (46), (47) may be written as:

$$x \sim N(\tilde{\xi}, \tilde{V}) \tag{48}$$

$$\tilde{B}'_\theta \tilde{\xi} = 0 \tag{49}$$

where:

$$\tilde{B}'_\theta = \begin{pmatrix} \alpha & 1 & 0 & . & \ldots & 0 & 0 & \gamma & \beta & 0 & . & \ldots & 0 & 0 \\ 0 & \alpha & 1 & 0 & \ldots & 0 & 0 & 0 & \gamma & \beta & 0 & \ldots & 0 & 0 \\ 0 & 0 & . & . & \ldots & \alpha & 1 & 0 & . & . & . & \ldots & \gamma & \beta \end{pmatrix} \tag{50}$$

$$(T-1) \times 2T$$

The global cut condition applied to z is then:

$$\begin{cases} \operatorname{Cov}(v_{yt} + \alpha v_{y\,t-1} + \beta v_{zt} + \alpha v_{z\,t-1}, v_{zt'}) = 0 \\ \forall t' = 1, \ldots T, \qquad t = 2, \ldots, T \end{cases} \tag{51}$$

If the errors process is stationary and if $\sigma_{zz}(k)$ and $\sigma_{yz}(k)$ represent the autocovariance function of v_z and the cross covariance function ($\sigma_{yz}(k)=\text{cov}(v_{yt}, v_{z,t+k})\, k\in\mathbb{Z}$), (51) is equivalent to:

$$\sigma_{yz}(k)+\alpha\sigma_{yz}(k-1)+\beta\sigma_{zz}(k)+\gamma\sigma_{zz}(k-1)=0. \tag{52}$$

If the global cut condition holds for any sample size, then (52) is true for any k and if $|\alpha|<1$, this assumption means that the $\sigma_{yz}(k)$ function is determined by the $\sigma_{zz}(k)$ function and by the parameters. Indeed, one has $\sigma_{yz}(k)=\sum_{i=0}^{\infty}\nu_i\sigma_{zz}(k-i)$ where the ν_i are the coefficients of the serie $(1+\alpha X)^{-1}(\beta+\gamma X)$.

If (51) is verified then (47) may be rewritten as:

$$\xi_{yt}^{z}+\alpha\xi_{yt-1}^{z}+\beta z_t+\gamma z_{t-1}=0.$$

3.2. Innovations

We now return to the general specification $x_t=\xi_t+v_t$ where $B'_\theta(L)\,\xi_t+C'_\theta(L)\,w_t=0$. w_t now incorporates the initial conditioning variables and the variables satisfying the global cut property.

This model will now be examined sequentially using the probability generating x_t given $x_{t-1}, x_{t-2}, \ldots$

If η_t denotes the conditional expectation of x_t given its history and V_t the conditional variance of x_t, the sampling process may be written as:

$$x_t\,|\,x_{t-1}, \ldots\, \lambda \sim N(\eta_t, V_t). \tag{53}$$

In general, the $\eta_t s$ are subject to linear constraints involving their whole past.

For example, if we consider the model for any $t\in\mathbb{Z}$ and if v_t is assumed to be stationary verifying an (infinite or finite) autoregressive representation $v_t=\Pi(L)\,v_t+\varepsilon_t$ [where ε_t is the innovation of v_t and $\Pi(L)$ is a matrix of series without constant term in the lag operator], the condition $B'_\theta(L)\,\xi_t+C'_\theta(L)\,w_t=0$ becomes

$$B'_\theta(L)\,(I-\Pi(L))^{-1}\,\eta_t-B'_\theta(L)\,(I-\Pi(L))^{-1}\,\Pi(L)\,x_t+C'_\theta(L)\,w_t=0. \tag{54}$$

Then the distribution of x_t given its history depends on the whole sequence (infinite or truncated at the initial conditions) of the incidental parameters. The following assumption avoids this difficulty.

The $u_t^\theta=B'_\theta(L)\,x_t+C'_\theta(L)\,w_t$ process is an *x-innovation* if u_t^θ is independent (given w and λ) from the past $(x_{t-1}, x_{t-2}, \ldots)$ of the x_t process.

Proposition 2 : The two conditions:

(i) $B'_\theta(L)\,\xi_t+C'_\theta(L)\,w_t=0$;

(ii) $B_\theta^{0\prime}\,\eta_t+B_\theta^{*\prime}(L)\,x_t+C'_\theta((L)\,w_t=0$

$[B_\theta(L)=B_\theta^0+B_\theta^*(L)$ where B_θ^0 is the matrix of constant terms of $B_\theta(L)]$ are equivalent if and only if the $u_{\theta t}$ process is an x-innovation.

The proof is given in the appendix.

Let us remark that the innovation property of u_t^θ implies that u_t^θ is independent of its own past but this last property is weaker then the innovation condition.

Example 2: We continue the example 1 without assuming the global cut condition. The x-innovation condition may be written as

$$(54)\qquad \begin{cases} \operatorname{Cov}(v_{yt}+\alpha v_{yt-1}+\beta v_{zt}+\gamma v_{zt-1},\ \rho v_{yt-k}+\mu v_{zt-k})=0 \\ \qquad\qquad \forall \rho,\ \mu R, \qquad \forall k \geqslant 1. \end{cases}$$

Using a stationarity hypothesis and if $\sigma_{yy}(k)$ denotes the autocorrelation function of v_y, (54) is equivalent to:

$$(55)\qquad \sigma_{yy}(k)+\alpha\sigma_{yy}(k-1)+\beta\sigma_{yz}(k)+\gamma\sigma_{yz}(k-1)=0$$

and

$$(56)\qquad \sigma_{yz}(-k)+\alpha\sigma_{yz}(-k+1)+\beta\delta_{zz}(k)+\gamma\sigma_{zz}(k-1)=0.$$

If (54) is satisfied, then (47) may be rewritten as

$$\eta_{yt}+\alpha y_{t-1}+\beta\eta_{zt}+\gamma z_{t-1}=0.$$

3.3. Sequential Cut

As in the paragraph 3.1, we first recall the general definition of a sequential cut. Let $x=(x_t)$ be a vector of observations and $l(x \mid w, \lambda)$ be its likelihood given some conditioning variables w and the parameters λ. $l(x \mid w, \lambda)$ is decomposed sequentially:

$$(57)\qquad l(x \mid w, \lambda)=\prod_t l(x_t \mid x_{t-1}, x_{t-2}, \ldots, w, \lambda)$$

x_t is now partionned in (y_t, z_t) and this partition realizes a *sequential cut* if there exists a partition of λ in $(\lambda_m^s, \lambda_c^s)$ such that:

(i) λ_c^s and λ_m^s are (sequentially) sufficient parameters for the model generating y_t given $z_t, x_{t-1}, \ldots, w, \lambda$ and for the model generating z_t given $x_{t-1}, x_{t-2}, \ldots, w, \lambda$, respectively. Equivalently:

$$(58)\qquad l(y_t \mid z_t, x_{t-1}, \ldots, w, \lambda)=l(y_t \mid z_t, x_{t-1}, \ldots w, \lambda_c^s)$$

$$(59)\qquad l(z_t \mid x_{t-1}, \ldots, w, \lambda)=l(z_t \mid x_{t-1}, \ldots, w, \lambda_m^s)$$

(ii) λ_m^s and λ_c^s are variation free (for a classical sequential cut) or *a priori* independent (for a bayesian sequential cut).

This property has an intuitive interpretation analogous to the interpretation of a global cut except that, in this case, the decomposition is realized at each step of the generating process. The main consequence is that the factorization derived from the cut property cannot be interpreted as a decomposition in a marginal and a conditionnal distribution. From (58) and (59) it follows that

$$(60)\quad l(x \mid w, \lambda)=\prod_t l(z_t \mid x_{t-1}, \ldots w, \lambda_m^s)\prod_t l(y_t \mid z_t, x_{t-1}, \ldots, w, \lambda_c^s).$$

The two factors of the product are not data densities even if each of them is sufficient for the inference on the related parameter. Note moreover that it could happen that the z_t process realizes both a global and a sequential cut with a different decomposition of λ.

In the linear model the sequential cut condition is useful essentially when the $u_{\theta t}$ are x-innovations because this last condition is required for a tractable sequential analysis.

PROPOSITION 3 : In the model (1) (2), let us assume that $u_{\theta t}$ are x-innovations. The partition $x_t' = (y_t', z_t')$ realizes a (classical) sequential cut if and only if $u_{\theta t}$ and z_t are independent given $x_{t-1}, \ldots, w, \lambda$.

The proof is identical to the proof of the proposition 1 using the sequential form of the model (53) and the expression the linear constraint by the form (ii) of the proposition 2.

Under the conditions of the proposition 3, the sequential model may be decomposed into the marginal model:

(61) $$z_t \mid x_{t-1}, \ldots w, \eta_{zt}, V_{zzt} \sim N(\eta_{zt}, V_{zzt})$$

and the conditional model

(62) $$y_t \mid z_t, x_{t-1}, \ldots, w, \theta, \eta_{yt}^z, V_{yy.zt} \sim N(\eta_{y,t}^z, V_{yy.zt})$$

in which the $\eta_{yt}^z = E(y_t \mid z_t, x_{t-1}, \ldots, w, \lambda)$ verifies:

$$B_\theta^{0\prime} \begin{pmatrix} \eta_{yt}^z \\ z_t \end{pmatrix} + B_\theta^{*\prime}(L)\, x_t + C_\theta'(L)\, w_t = 0.$$

The two sets of parameters $\lambda_m^s = (\eta_{zt}), (V_{zzt})$ and $\lambda_c^s = (\theta, \eta_{yt}^z, (V_{yy.zt})_t)$ are variation free if as in 1 it is implicitely assumed that V_{zzt} (resp. $V_{yy.zt}$) is not related to the parameters of the conditional (resp. marginal) model.

Example 3: We start from the example 1 and we assume that the innovation property is verified. The sequential model is defined by:

(63) $$\begin{pmatrix} y_t \\ z_t \end{pmatrix} \Bigg| x_{t-1}, \ldots, \lambda \sim N\left[\begin{pmatrix} \eta_{yt} \\ \eta_{zt} \end{pmatrix}, V\right] \qquad V = \begin{pmatrix} V_{yy} & V_{yz} \\ V_{zy} & V_{zz} \end{pmatrix}$$

under a stationarity assumption. The expectation satisfies:

(64) $$\eta_{yt} + \alpha y_{t-1} + \beta \eta_{zt} + \gamma z_{t-1} = 0$$

The sequential cut property in this example reduces to the independence between z_t and $y_t + \beta z_t$ which is equivalent to $V_{yz} + \beta V_{zz} = 0$.

If this condition is verified the conditional sequential model becomes

(65) $$y_t \mid z_t, x_{t-1}, \ldots \sim N(\eta_{yt}^z, V_{yy.z})$$

where

$$V_{yy.z} = V_{yy} - \frac{V_{yz}^2}{V_{zz}}$$

and

(66) $$\eta_{yt} + \alpha y_{t-1} + \beta z_t + \gamma z_{t-1} = 0.$$

Thos model does not involved any incidental parameter and reduces to a dynamic regression model.

3.4. Non Causality

A non causality assumption may be introduced for two kinds of reasons. It could be suggested by economic theory either as an hypothesis to be specified at the stage of model building or as an hypothesis to be tested at the stage of inference. The second reason is motivated by the relation between the two differents cuts wich are equivalent under a non causality assumption. Essentially if $x_t = (y_t, z_t)$ a global cut is equivalent to a sequential cut if y does not cause z. This result requires additional technical assumptions and is correctly stated in FLORENS *et al.* [1980], FLORENS-MOUCHART [1985].

The GRANGER'S [1969] *condition of non causality* is now well known. The y_t process does not cause the z_t process if y_t is independent of $z_{t-1}, z_{t-2}, \ldots$ given $y_{t-1}, y_{t-2}, \ldots$ [and given (λ, w)].

In the linear dynamic model this condition is not related to the equations constraining the expectations of the variables but is only a property of the distribution of the residuals.

PROPOSITION 4 : In the model (1), (2), the y_t process does not cause the z_t process if and only if the v_{yt} process does not cause the v_{zt} process given λ and w where $v_t = x_t - \xi_t$ is decomposed in $v_t = (v_{yt}, v_{zt})$.

A precise proof of this proposition in given in FLORENS *et al.* [1986]. The result is intuitive because the non causality is a property of the sampling process i. e. is conditional on the parameters. Thus the independence is not affected by a transformation of the variables into the deviations around their expectations which are considered as parameters.

We conclude this section by an example also given in FLORENS *et al.* [1986] which illustrates that under an innovation property the noncausality hypothesis links the parameters of the equations and the parameters of the residual process.

Example 4 : Consider the following bivariate model:

$$\begin{pmatrix} y_t \\ z_t \end{pmatrix} = \begin{pmatrix} \xi_{yt} \\ \xi_{zt} \end{pmatrix} + v_t \tag{67}$$

$$\xi_{yt} + \alpha \xi_{zt} + \beta \xi_{zt} + \gamma \xi_{zt-1} = 0 \tag{68}$$

where v_t is a stationary AR (1) process represented as

$$v_t = \mathrm{R}\, v_{t-1} + e_t, \qquad \mathrm{R} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{69}$$

where $e_t \sim \mathrm{IN}(0, \Phi)$ is the innovation of v_t.

If $b=0$ then z does not cause y while if $c=0$ y does not cause z. By construction of the model, it can be verified that u_t^θ is an x-innovation if and only if $u_{\theta t}$ is independent of v_{t-1} in the sampling process. Under the stationarity assumption the covariance matrix between v_{t-1} and v_t is equal to:

$$\begin{pmatrix} \mathrm{V} & \mathrm{VR}' \\ \mathrm{RV} & \mathrm{V} \end{pmatrix} \tag{70}$$

where V is the solution of the system $V = \Phi + RVR'$. Therefore as V is non singular, u_t^θ is an x-innovation if and only if

$$(1 \quad \alpha) R + (\beta \quad \gamma) = 0. \tag{71}$$

In such a case and if, furthermore $a = 0$, the condition $\beta = 0$ is equivalent to the condition $c = 0$ which means that y does not cauze z (under $\alpha \neq 0$). Thus under (71) an hypothesis of noncausality may be written as a restriction on $\theta = (\alpha, \beta, \gamma)$. Note also that the condition for a sequential cut is

$$(1 \quad \alpha)\, \varphi \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 0. \tag{72}$$

3.5 *Reduction of a Model Relatively to a Subset of Equations*

In the preceding sections we have analysed the problem of the reduction of a dynamic linear model without losing any information about all the parameters of the equations (2).

Often the parameters of interest are concentrated in a subset of equations and a reduction may be considered as admissible if no information is lost on the parameters of this subset of equations. We illustrate this remark on an example close to the one given at the end of section 2.

Let us start with the model generating $x_t = (y_t, z_t)$ given (w_t)

$$\begin{cases} y_t = \xi_{yt} + v_{yt} \\ z_{zt} = \xi_{zt} + v_{zt} \end{cases} \tag{73}$$

$$\begin{cases} \xi_{yt} + \beta_1 \xi_{yt-1} + \beta_2 \xi_{zt} = 0 \\ \xi_{zt} - \delta w_t = 0 \end{cases} \tag{74}$$

in which we operate the transformation:

$$\begin{aligned} u_t^\theta &= y_t + \beta_1 y_{t-1} + \beta_2 z_t \\ z_t &= \delta w_t + \mu z_{t-1} + \lambda_{+1} u_{t+1}^\theta + \lambda_0 u_t^\theta + \lambda_{-1} u_{t-1}^\theta + \varepsilon_t \end{aligned} \tag{75}$$

in which the second relation describes the process generating z_t given its past, and the whole trajectory of u_t and of w_t. ε_t is a white noise independent by construction from $z_t, z_{t-1}, \ldots, (w_t)$ and (u_t^θ).

The specification is achieved by the distribution of the u_θ process which is assumed to be an AR(1) process verifying $u_t^\theta = \rho u_{t-1}^\theta + e_t$. Here y_0 and z_0 are included in the conditioning process w_t and are treated as non random.

In the spirit of the section 2, the parameters of interest are β_1, and β_2 and we want to apply the conditional independence conditions presented above in order to reduce the model by conditioning without losing any information on β_1 and β_2.

Since the modelisation has been done by defining separately the u_t^θ process and the z_t process given the u_t^θ process, the simplest property is the global cut condition. If $\lambda_{+1} = \lambda_0 = \lambda_{-1} = 0$, z_t realizes a global cut relatively on the parameters of the first equation. The marginal process on z_t does not bring any information on β_1 and β_2 and the inference can be concentrated on the y_t process given z_t which is characterized by the distribution of u_θ and the relation $\xi^z_{yt} + \beta_1 \xi^z_{yt-1} + \beta_2 z_t = 0$ where ξ^z_{yt} is the sampling expectation of y_t given $(z_t)_t$.

The x-innovation property can be easily characterized by a double condition. u_θ is an innovation for the (y_t, z_t) process if u_θ is a white noise ($\rho = 0$) and if $\lambda_{+1} = 0$. In this case the model can be written sequentially. Let (η_{yt}, η_{zt}) be the conditional expectations of (y_t, z_t) given x_{t-1}, $x_{t-2} \ldots$ These parameters are related by the relations

$$(76) \qquad \begin{cases} \eta_{yt} + \beta_1 y_{t-1} + \beta_2 \eta_{zt} = 0 \\ \eta_{zt} = \delta w_t + \mu z_{t-1} + \lambda_{-1} u^\theta_{t-1}. \end{cases}$$

The noncausality relation may be analysed through the distribution of z_t given $x_{t-1}, x_{t-2} \ldots$ We simplify our approach by assuming that u_θ is an x-innovation. In this case z_t given the x-past has an expectation equal to

$$\delta w_t + (\mu + \lambda_{-1} \beta) z_{t-1} + \lambda_{-1} y_{t-1} + \lambda_{-1} \beta_1 y_{t-2}$$

and y_t does not cause z_t if $\lambda_{-1} = 0$.

In the model presented here a sequential cut is not easily characterized by a minimal set of conditions, even if the innovation condition is realized, because a lagged value of u^θ_t appears in the second equation. However a sequential cut may be obtained as an implication of a global cut and a non causality property.

Proof of Proposition 1

This proposition is based upon the following lemma.

LEMMA : Let E be a linear manifold of R^{n+k} defined by :

$$E=\{(y, z)/y\in\mathbb{R}^n, z\in\mathbb{R}^k \quad By+Cz+a=0 \quad B\, p\times n$$
$$rk(B)=p, C \quad p\times k, a\in\mathbb{R}^p\}.$$

The following two properties are equivalent:

(i) $\exists$ F linear manifold of R^n and G linear manifold of R^k such that $(y, z)\in E \Leftrightarrow y\in F$ and $z\in G$ (i. e.: $E=F\times G$).

(ii) $C=0$.

Proof: (ii) implies trivially (i) taking $G=\mathbb{R}^k$ and

$$F=\{y\in\mathbb{R}^n/By+a=0\}.$$

Let us now assume (i) to be true. Let us first remark that one has necessarily $G=\mathbb{R}^k$. As $rk(B)=p$, dim $E=(n+k)-p$. Let $z_0\in G$. F can be written in particular as $\{y\in\mathbb{R}^n \mid By+Cz_0+a=0\}$ which has a dimension equal to $n-p$. (i) implies that dim E = dim F + dim G so that dim $G=k$ and $G=R^k$. Finally we remark that for any $z_1, z_2\in\mathbb{R}^k$

$$F=\{y \mid By+Cz_1+a=0\}=\{y/By+Cz_2+a=0\}.$$

Thus $C(z_1-z_2)=0$ and $C=0$. □

Proof of Proposition 1

Let us consider the marginal and the conditional sampling distributions:

$$z \mid \lambda, w \sim N(\tilde{\xi}_z, \tilde{V}_{zz})$$
$$y \mid \lambda, z, w \sim N(\tilde{\eta}_y, \tilde{V}_{yy\,.\,z})$$

where

$$\tilde{\eta}_y=\tilde{\xi}_y+\tilde{V}_{yz}\tilde{V}_{zz}^{-1}(z-\xi_z)$$

and

$$\tilde{V}_{yy\,.\,z}=\tilde{V}_{yy}-\tilde{V}_{yz}\tilde{V}_{zz}^{-1}\tilde{V}_{zy}$$

$\tilde{\xi}_z$ and $\tilde{\eta}_y$ are related by

$$\tilde{B}_\theta^{y'}\tilde{\eta}_y+(\tilde{B}_\theta^{y'}\tilde{V}_{zz}^{-1}+\tilde{B}_\theta^{z'})\xi_z-\tilde{B}_\theta^{y'}\tilde{V}_{yz}\tilde{V}_{zz}^{-1}z+\tilde{C}_\theta' w=0.$$

As $\tilde{V}_{zz}$ and $\tilde{V}_{yy.z}$ can be arbitrary PDS matrices, the parameters of the marginal and the conditional models are variation free if and only if, using the preceding lemma,

$$\tilde{B}_{\theta}^{y'} \tilde{V}_{yz} \tilde{V}_{zz}^{-1} + \tilde{B}_{\theta}^{z'} = 0.$$

This condition is equivalent to $\tilde{B}_{\theta}^{y'} \tilde{V}_{yz} + \tilde{B}_{\theta}^{z} \tilde{V}_{zz} = 0$ which characterizes the nullity of the covariance between u_{θ} and z and hence the independence in the normal framework.

Under this hypothesis the parameters of the marginal model are arbitrary elements of $R^k \times \mathscr{C}_{\tilde{k}} (z \in \mathbb{R}^k)$ ($\mathscr{C}_{\tilde{k}}$ is the cone of PDS matrices $\tilde{k} \times \tilde{k}$).

The parameters $(\tilde{\eta}_y, \tilde{V}_{yy.z})$ of the conditional model are constrained to:

$$\tilde{B}_{\theta}' \begin{pmatrix} \tilde{\eta}_y \\ z \end{pmatrix} + \tilde{C}_{\theta}' w = 0, \qquad \tilde{V}_{yy.z} \in \mathscr{C}_{\tilde{p}} \qquad \square$$

Proof of Proposition 2

Let us first remark (ii) implies (i) without any assumption by taking the expectation of (ii).

(i) is equivalent to $B_{\theta}'(L) x_t + C_{\theta}'(L) w_t - B_{\theta}'(L) v_t = 0$ and as $B_{\theta}(L) v_t = u_t^{\theta}$ (i) is equivalent to (iii)

$$B_{\theta}'(L) x_t + C_{\theta}'(L) w_t - u_t^{\theta} = 0.$$

If u_t^{θ} is an innovation $E(u_t^{\theta} | x_{t-1}, \ldots, w,) = 0$ and taking the conditional expectation of (iii) given the history of x_t and (w, λ) gives (ii).

If (i) and (ii) are equivalent if follows that $E(u_{\theta t} | x_{t-1}, \ldots, w, \lambda) = 0$ which implies in the normal processes the innovation property. $\square$

• References

De Groot, M. H. (1970). – *Optimal Statistical Decisions*, MacGraw Hill, New York.

Deistler, M. (1986). – "Identifiability and Causality in Linear Dynamic Erros-in-Variables Systems". J. P. Florens (ed.) *Asymptotic Theory for non i.i.d. Processes*" Publication des Facultés Universitaires Saint-Louis, Bruxelles.

Engle, R. F., Hendry, D. F. and Richard, J. F. (1983). – "Exogeneity", *Econometrica*, 51, (2), p. 277-304.

Florens, J. P. and Mouchart, M. (1985). – "Conditioning in Dynamic Models", *Journal of Times Series* 6, (1), p. 15-35.

Florens, J. P. and Mouchard, M. (1986). – « Exhaustivité, ancillarité et identification en statistique bayésienne », *Annale d'Économie et de Statistique*, 4, p. 63-93.

Florens, J. P., Mouchart, M. and Richard, J. F., (1974). – "Bayesian Influence in Error-in-Variables Models", *Journal of Multivariate Analysis*, 4, p. 419-452.

Florens J. P., Mouchard, M. and Richard, J. F. (1976). – "Likelihood Analysis of Linear Models", *discussion paper* 7619, CORE, Université catholique de Louvain.

Florens, J. P., Mouchart, M. and Richard, J. F. (1979). – "Specification and Influence in Linear Models", *COREDP* 7943, Université Catholique de Louvain.

Florens, J. P., Mouchart, M. and Richard, J. F. (1986). – "Structural Time Series Modelling: Bayesian Approach", *Applied Mathematics and Computation*, 20 (3-4), p. 365-400.

Florens, J. P., Mouchard, M. and Rolin, J. M. (1980). – « Réductions dans les expériences bayésiennes séquentielles », Cahiers du *Centre d'Études de Recherche Opérationnelle*, 22 (3-4), p. 353-362.

Granger, C. W. J. (1969). – "Investigating Causal Relations by Econometric Models and Cross Spectral Methods", *Econometrica*, 37, p. 424-438.

Griliches, Z. (1974). – "Errors in Variables and Other Unobservables" *Econometrica*, 42, (6) p. 971-998.

Holly, A. (1983). – "A Simple Procedure for Testing wether a Subset of Endogeneous Variables is Independent of the Disturbance Term in a Structural Equation", *Cahiers de Recherches Economiques*, n° 8306, Université de Lausanne.

Holly, A. and Sargan, J. D., (1982). – "Testing for Exogeneity within a Limited Information Framework". *Cahiers de Recherches Economiques*, n° 8204, Université de Lausanne.

Richard, J. F. (1984). – "Classical and Bayesian Inference in Incomplete Simultaneous Equations Models", D. F. Hendry, and Wallis K. F. (eds.), *Econometrics and quantitative Economics*, Basic Blackwell, Oxford, p. 61-102.