U.S. Senate, Commitee on the Judiciary, subcommittee on Antitrust and Monopoly (1967). *International Aspects of Antitrust, 1967 : Review of the Webb-Pomerene Act of 1918*, Hearings pursuant to S. Res. 26, Cong., 1 sess. Washington : Government Printing Office.

ICAS A.G. et DEUTSCH A. – The Paradox of Employment Creation through mport Subsidies. *Economic Journal*, vol. 74, Mars, pp. 228-30, 1964.

FORMULATION GÉNÉRALE ET ESTIMATION DE MODÈLES MULTIDIMENSIONNELS TEMPORELS A FACTEURS EXPLICATIFS NON OBSERVABLES

par

Robert F. ENGLE et Mark WATSON

Université de Californie à San Diego

Encore peu utilisés en économétrie sous leur forme la plus générale, les modèles multidimensionnels temporels à facteurs explicatifs non observables relèvent directement de la «modélisation dans l'espace des états» bien connue des ingénieurs. On montre dans ce texte comment l'utilisation du filtre de Kalman et de la méthode des scores permet d'estimer par le maximum de vraisemblance les paramètres inconnus de ces modèles, y compris éventuellement les valeurs prises par les facteurs non observables. On indique également comment effectuer des tests de diagnostic (ou tests de spécification). On donne enfin les résultats d'une application de ces diverses techniques à l'estimation de la composante locale du salaire à Los Angelès.

1. INTRODUCTION

La plupart des économistes jugent fructueux d'utiliser des variables non observables pour décrire les phénomènes économiques. Ainsi, il y a longtemps que l'on traite les effets saisonniers ou les erreurs de mesure comme des composantes inobservables à extraire des séries statistiques. Dans d'autres modèles, on considère les fluctuations conjoncturelles ou les cycles longs comme des variables non observables qui déterminent indirectement l'évolution de séries observables : l'application la plus réussie de cette approche est la théorie du revenu permanent, où ce concept, bien que non mesurable, est utilisé pour expliquer les régularités observées dans diverses données économiques. Les modèles macroéconomiques récents sont truffés de variables

telles que les anticipations, le taux d'intérêt réel ou le taux «naturel» de chômage, variables qui ne sont pas observables mais qui, on présume, aident à expliquer les processus générateurs des données observées. En économie du travail, le talent, le «punch», l'hérédité sont traités comme des facteurs non observables qui influent sur l'éducation et le revenu.

Il existe de nombreuses possibilités d'extension de tels modèles. Dans tous les cas, on formule le modèle statistique sans se préoccuper de l'absence de données pour les variables non observées. On en déduit la loi de probabilité jointe des variables observables, qui peut alors servir de fonction de vraisemblance pour estimer les paramètres inconnus. Dans certains cas – et on en verra un exemple dans ce texte –, il est possible d'estimer aussi les valeurs des variables non observées.

Bon nombre des modèles auxquels il vient d'être fait allusion n'ont pas utilisé – ou du moins n'ont pas complètement utilisé – les restrictions a priori qui peuvent résulter d'une spécification détaillée des composantes non observées. Et même lorsqu'elles sont mises en œuvre de façon pleinement efficace, les procédures d'estimation sont trop compliquées et trop spécialisées pour qu'on puisse envisager facilement de modifier, même légèrement, le modèle initialement retenu et d'effectuer des tests de spécification.

En se fondant sur les modèles dans l'espace des états qu'utilisent les ingénieurs, on donne dans ce texte une méthode générale pour formuler et estimer les modèles à composantes non observées. Dans le paragraphe 2, on présente le modèle fondamental et on met en lumière sa généralité. Dans le paragraphe 3, on établit l'algorithme récursif de Kalman et la fonction de vraisemblance associée au modèle, tandis que dans le paragraphe 4, on étudie la maximisation de cette fonction par la méthode des scores. On présente plusieurs tests de spécification dans le paragraphe 5, et on donne enfin un exemple d'application dans le paragraphe 6.

2. FORMULATION GÉNÉRALE DU MODÈLE

Tous les modèles évoqués ci-dessus sont des cas particuliers du «modèle dans l'espace des états» qu'utilisent les ingénieurs pour représenter divers processus physiques. En fait, on verra ci-dessous qu'une large gamme de modèles utilisés en économétrie peuvent être considérés comme des modèles dans l'espace des états. On trouve dans Mehra [26] une introduction à ces modèles et une comparaison entre les applications qui en sont faites par les économètres et par les ingénieurs. L'avantage qu'il y a à considérer les modèles sous cet angle est que l'on dispose alors de concepts de solution très généraux fondés sur le principe de vraisemblance et sur l'algorithme du filtre de Kalman.

Un modèle dans l'espace des états comporte deux ensembles d'équations : des équations de «transition» ou «d'évolution», et des équations de «mesure». Les équations de transition décrivent l'évolution d'un vecteur x_t de J caractéristiques physiques en réponse à l'évolution d'un vecteur z_t de K variables exogènes (ou endogènes retardées) et d'un vecteur v_t de J perturbations. Ce vecteur d'états x_t est inobservable et représente donc les composantes non observées qu'on cherche à isoler. Les équations de «mesure» décrivent les relations entre le vecteur x_t des états inobservables et un vecteur y_t de P mesures. Le vecteur z_t des variables prédéterminées et un autre vecteur de perturbations, e_t, peuvent également apparaître dans les équations de mesure.

Le modèle peut donc être spécifié ainsi :

$$\underset{(J,1)}{x_t} = \underset{(J,J)}{\Phi_t}\,\underset{(J,1)}{x_{t-1}} + \underset{(J,K)}{C_t}\,\underset{(K,1)}{z_t} + \underset{(J,1)}{v_t} \tag{1}$$

$$\underset{(P,1)}{y_t} = \underset{(P,J)}{A_t}\,\underset{(J,1)}{x_t} + \underset{(P,K)}{B_t}\,\underset{(K,1)}{z_t} + \underset{(P,1)}{e_t} \tag{2}$$

$$\begin{bmatrix} v_t \\ e_t \end{bmatrix} \sim \mathcal{N}\left[0, \begin{bmatrix} \underset{(J,J)}{Q_t} & 0 \\ 0 & \underset{(P,P)}{R_t} \end{bmatrix}\right] \tag{3}$$

Pour les applications effectuées dans ce texte, les paramètres et la matrice des variances-covariances sont constants dans le temps, et n'auront donc pas d'indice temporel. L'algorithme de Kalman et certaines des applications économiques qui seront décrites ci-dessous nécessitent toutefois des indices temporels ; aussi le modèle fondamental est-il présenté ici sous sa forme la plus générale.

En économie, on a largement utilisé ce modèle dans le cas où il ne comporte qu'une seule équation de mesure. Si $A = 0$ ou $Q = 0$ ce modèle est un modèle de régression linéaire, tandis que c'est un modèle ARIMA si $B = 0$ et $C = 0$. Hannan [23] a montré qu'il y a une correspondance biunivoque entre les modèles dans l'espace des états et les modèles ARIMA. Pour écrire comme des processus du premier ordre du type de l'équation (1) les modèles ARIMA d'ordre supérieur, il est nécessaire d'augmenter la dimension du vecteur des états (cf. par exemple Chow [6]). De la même manière, si $B \neq 0$, la perturbation devient $Ax_t + e_t$, qui est un processus d'erreurs ARIMA ; le modèle est alors un modèle de régression avec des perturbations ARIMA, comme cela a été noté par Harvey et Phillips [25].

Nerlove [28], Pagan [29] et Engle [11] ont développé des applications où les x_t sont traités comme des composantes inobservables. Dans ces modèles les termes d'erreur sont décomposés en plusieurs composantes dont

les propriétés temporelles sont différentes. Ainsi Pagan [29] a décomposé la propension moyenne à la consommation en une propension permanente et une propension transitoire, pour chacune desquelles il a admis des hypothèses temporelles différentes ; Engle [11] a utilisé la même technique pour l'ajustement saisonnier, les composantes saisonnières et non saisonnières étant par définition supposées obéir à des processus temporels différents.

On obtient un ensemble très riche d'applications lorsqu'on considère les x_t comme des coefficients de regression susceptibles de varier dans le temps, la matrice A_t étant alors la matrice des observations sur les variables exogènes à la période t. Dans cette interprétation, les z_t pourraient être les variables dont les coefficients ne changent pas au cours du temps. En faisant $C = 0$ et $Q = 0$, et en prenant pour Φ la matrice identité, on retombe sur le cas de paramètres constants dans le temps. Les résidus de ce modèle (qui seront ci-dessous appelés «innovations») sont appelés résidus récursifs et servent à tester si les paramètres sont ou non variables (cf. Brown, Durbin et Evans [4] Harvey et Collier [24] et Garbade [16]). D'autre part, si $Q \neq 0$ et $\Phi = 0$, le modèle devient un modèle à coefficients aléatoires. Cooley et Prescott [7], [8] se sont intéressés à un modèle de regression adaptative dans lequel le terme constant et certains coefficients comportent une composante permanente et une composante transitoire variables dans le temps ; autrement dit, les coefficients suivent un modèle ARIMA (0, 1, 1). Pagan [30] et Rosenberg [31] supposent que les coefficients suivent un modèle ARIMA général.

A cette multiplicité d'applications du modèle dans l'espace des états avec une seule équation de mesure s'oppose l'absence apparente d'applications du modèle complet aux séries temporelles dans le cas $P > 1$. L'analyse instantanée de tels modèles est pourtant très riche. Lorsque $\Phi = 0$, les observations successives sont, d'après (3), indépendantes aussi longtemps que z ne comporte que des variables strictement exogènes. Lorsqu'à la fois $\Phi = 0$, $B = 0$, $C = 0$ et que Q est diagonale, le modèle se ramène au modèle classique d'analyse factorielle ou d'indicateurs multiples avec P mesures (ou indicateurs) et J facteurs, les poids des facteurs étant donnés par A. Si $B \neq 0$, Q non diagonale, mais $\Phi = 0$ et $C = 0$, le modèle n'est autre qu'un modèle de régression multidimensionnelle (empilement de régressions sur les mêmes variables explicatives), ou encore la forme réduite d'un modèle à équations simultanées. La structure d'analyse factorielle de la matrice des variances-covariances des perturbations pourrait entrainer des restrictions sur les covariances entre équations. Supposer de plus, $C \neq 0$ revient à admettre l'existence d'un ensemble de variables causales pour les composantes non observées. Les techniques initialement présentées par Zellner [36] et Goldberger [19] et plus tard développées par Goldberger [20] sont adaptées au traitement de ce modèle, baptisé MIMIC par ces auteurs (modèle à multiples indicateurs et multiples causes).

Lorsque $\Phi = 0$ et $P > 1$, le modèle dans l'espace des états représente un processus temporel multivarié. Lorsque $C = 0$ et $B = 0$, c'est un processus ARIMA multivarié général dont la discussion a été présentée notamment par Granger et Newbold [21] et l'estimation par Wallis [35]. Lorsque $B = 0$, le modèle peut l'interpréter comme la forme réduite d'un système d'équations simultanées à perturbations auto-corrélées.

Bien que Geweke [17], Brillinger [3] et Sargent-Sims [32] aient proposé plusieurs procédures dans le domaine fréquentiel pour identifier les composantes non observées de séries temporelles multivariées, les auteurs du présent texte ne connaissent qu'une seule approche analogue dans le domaine temporel. Sargent et Sims [32] et Sims [33] ont introduit un modèle à «indice observable» qui définit la composante non observée comme une combinaison linéaire non aléatoire de variables endogènes retardées. Le vecteur z inclut alors des valeurs retardées du vecteur y et les perturbations associées sont de variance nulle, de sorte que $Q = 0$.

3. LA FONCTION DE VRAISEMBLANCE

Le filtre de Kalman et la fonction de vraisemblance des paramètres s'obtiennent facilement à l'aide des propriétés classiques de la loi normale multidimensionnelle. Ainsi, il est bien connu (cf. Morrison [27] par exemple) que si

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim \mathcal{N}\left[\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right]$$

alors la distribution de X_1 conditionnée par X_2 est normale multidimensionnelle avec pour espérance le vecteur

$$\mu_1 + \Sigma_{12}\,\Sigma_{22}^{-1}\,(X_2 - \mu_2) \tag{4}$$

et pour matrice de variance-covariance :

$$\Sigma_{11} - \Sigma_{12}\,\Sigma_{22}^{-1}\,\Sigma_{21} \tag{5}$$

L'algorithme de Kalman est une méthode récursive pour calculer $E(x_t \mid Y_t)$, où $Y_t = [y_t, y_{t-1}, \ldots, y_1]$. Notons d'abord que, puisque la loi de probabilité jointe de x_t et Y_t est normale, la densité de $x_t \mid Y_t$ est normale multidimensionnelle ; nous pouvons la noter $f(x_t \mid y_t, Y_{t-1})$. Puis

$$E(x_t \mid Y_t) = E(x_t \mid y_t, Y_{t-1}),$$

et on peut obtenir cette dernière espérance en utilisant l'expression de la loi jointe de $(x_t, y_t \mid Y_{t-1})$. En adoptant les notations suivantes :

$$E(x_t \mid Y_{t-1}) := x_{t|t-1} \quad E(y_t \mid Y_{t-1}) := y_{t|t-1}$$

$$\text{var}\ (x_t \mid Y_{t-1}) := p_{t|t-1} \quad \text{var}\ (y_t \mid Y_{t-1}) := H_t$$

on déduit facilement des équations (1) et (2) les relations suivantes :

$$x_{t|t-1} = \Phi_t x_{t-1|t-1} + C_t z_t \tag{6}$$

$$p_{t|t-1} = \Phi_t p_{t-1|t-1} \Phi_t' + Q_t \tag{7}$$

$$y_{t|t-1} = A_t x_{t|t-1} + B_t z_t \tag{8}$$

$$H_t = A_t p_{t|t-1} A_t' + R_t \tag{9}$$

ainsi que

$$\text{cov}\,(x_{t|t-1}, y_{t|t-1}) = p_{t|t-1} A_t'$$

Ces équations fournissent l'espérance et la matrice de variance-covariance de la loi jointe de (y_t, x_t) conditionnée par Y_{t-1}. On peut alors utiliser (4) et (5) pour obtenir :

$$x_{t|t} = x_{t|t-1} + p_{t|t-1} A_t' H_t^{-1} (y_t - y_{t|t-1}) \tag{10}$$

et

$$p_{t|t} = p_{t|t-1} + p_{t|t-1} A_t' H_t^{-1} A_t p_{t|t-1} \tag{11}$$

Ces équations (6) à (11) définissent le filtre de Kalman, Cette technique d'estimation nécessite une valeur de $x_{0|0}$ et $P_{0|0}$ pour que la récurrence puisse démarrer. Très souvent, ces valeurs initiales résultent directement du modèle.

Dans le cas le plus simple $C_t = 0$, $\Phi_t = \Phi$ pour tout t, et $|\lambda_i(\Phi)| < 1$ pour tout i, $\lambda_i(\Phi)$ étant la i-ème valeur propre de Φ. Dans ces conditions, le processus x_t est stationnaire et les valeurs initiales sont simplement l'espérance et la variance du processus. Si $C \neq 0$ ou $|\lambda_i(\Phi)| > 1$ pour au moins un i, alors le filtre peut être amorcé en choisissant une valeur a priori telle que $x_{0|0} = 0$ et $P_{0|0} = \delta I$ où δ est un grand nombre quelconque. Plutôt que supposer les valeurs initiales déterminées par le même processus stochastique que les autres données, il peut être intéressant de les traiter comme des paramètres non aléatoires mais inconnus qu'il convient d'estimer. On présente dans le paragraphe suivant une solution à ce problème d'estimation en même temps que l'estimation par le maximum de vraisemblance d'autres paramètres.

Dans certains cas, par exemple le modèle DYMIMIC ou les modèles à composantes non observées, les paramètres du modèle ne varient pas dans le temps et n'ont donc pas d'indice t. Le modèle dans l'espace des états est

alors qualifié d'invariant. Lorsqu'un modèle est invariant et qu'il satisfait à certaines autres conditions, le filtre de Kalman est considérablement simplifié. En particulier, si $C_t = 0$ et $|\lambda_i(\Phi)| < 1$ pour tout i, alors on peut établir que (cf. Anderson et Moore [1])

$$\lim_{t \to \infty} P_{t|t} = \bar{P}$$

de sorte que $P_{t|t}$, H_t et $P_{t|t-1}$ tendent vers des constantes. En ce cas, les équations (7), (9) et (11) n'ont pas à être calculées à chaque période ; on se contente de remplacer les valeurs de $P_{t|t-1}$, H_t et $P_{t|t}$ par les valeurs qu'elles prendraient en régime stationnaire.

Les équations du régime stationnaire sont non-linéaires et difficiles à résoudre analytiquement ; mais on peut facilement les résoudre en pratique : il suffit d'utiliser les équations (6) à (11) pendant quelques périodes et de vérifier si $P_{t|t}$ a convergé. Si oui (ce qui se produit d'habitude aux alentours de $t = 10$, d'après notre expérience), on utilise la valeur ainsi trouvée pour le reste des récurrences.

On peut également utiliser les équations du filtre de Kalman pour obtenir une fonction de vraisemblance permettant d'estimer les paramètres inconnus du modèle. La fonction de vraisemblance est fondée sur les innovations du filtre, c'est-à-dire les erreurs de prévision à échéance d'une période, $y_t - y_{t|t-1}$. Appelons η_t le vecteur d'innovations,

On obtient alors très facilement la fonction de vraisemblance. Remarquons d'abord que la loi jointe de $(y_0, \ldots, y_T)$ peut s'écrire comme le produit de toutes les lois conditionnelles, de sorte que la log-vraisemblance peut s'écrire :

$$L = \text{Log}\, f(y_0) + \sum_{t=1}^{T} \text{Log}\,[f(y_t \mid Y_{t-1})]$$

Chacune des densités conditionnelles est normale, son espérance et sa variance étant données par (8) et (9). Si l'on dispose d'une distribution a priori non dégénérée pour y_0, la log vraisemblance peut s'écrire :

$$L = \text{constante} - \frac{1}{2} \sum_{t=1}^{T} (\text{Log}\,|H_t| + \eta_t' H_t^{-1} \eta_t) := \sum_t L_t \tag{12}$$

Si la distribution de la valeur initiale y_0 dépend des paramètres inconnus, alors on peut l'incorporer dans la fonction de vraisemblance et l'estimer selon la procédure déjà décrite. Asymptotiquement, il n'y a évidemment aucune différence.

4. MAXIMISATION DE LA VRAISEMBLANCE PAR LA MÉTHODE DES SCORES

A partir des données et de l'expression de la fonction de vraisemblance, il est en principe simple de maximiser cette dernière par rapport aux paramètres inconnus. En pratique, malheureusement, cette maximisation ne s'effectue pas si facilement du fait qu'il y a généralement un grand nombre de paramètres inconnus et que chaque évaluation numérique de la vraisemblance nécessite de nombreux calculs. C'est pourquoi il est important d'employer un algorithme qui n'utilise que les dérivées premières et qui, à partir de valeurs initiales convergentes, permette d'obtenir des estimateurs asymptotiquement pleinement efficaces en une seule étape, et efficaces au second ordre en deux étapes. Cela exclut immédiatement bon nombre d'algorithmes. L'algorithme classique de Davidson-Fletcher-Powell (DFP) n'utilise bien que les dérivées premières, mais il n'est pas pleinement efficace en une seule étape : DFP utilise pour matrice initiale la matrice identité, alors qu'il faut un estimateur convergent de la matrice d'information pour parvenir à l'efficacité en une étape. Aussi la méthode des scores paraît-elle ici une méthode intéressante ; et c'est une généralisation de l'approche de Pagan [30] que nous avons trouvée la meilleure.

Commençons par rassembler tous les paramètres inconnus en un vecteur θ. La procédure de calcul itérative consiste à obtenir $\mathcal{I}_k^{-1}$, un estimateur de l'inverse de la matrice d'information fondé sur l'estimateur θ_k de θ obtenu à la $k^{\text{ième}}$ itération, à choisir un scalaire λ_k définissant la longueur du $k^{\text{ième}}$ pas, et à obtenir un nouvel estimateur θ_{k+1} d'après la formule

$$\theta_{k+1} = \theta_k + \lambda_k \mathcal{I}_k^{-1} \left(\frac{\partial L}{\partial \theta}\right)_{\theta_k}.$$

Les dérivées nécessaires peuvent être calculées numériquement en effectuant autant de passes à travers le filtre de Kalman qu'il y a de paramètres inconnus indépendants dans θ.

Pagan [30] a obtenu une expression générale de la matrice d'information dans le cas $P = 1$. Nous allons présenter ici une formule analogue pour le cas $P > 1$ et corriger une erreur dans Gupta et Mehra [22]. Pour cela, nous utiliserons les formules bien connues de dérivation d'une matrice symétrique régulière :

$$\frac{\partial |B|}{\partial x} = |B| \operatorname{Tr}\left(B^{-1} \frac{\partial B}{\partial x}\right) \tag{13}$$

$$\frac{\partial B^{-1}}{\partial x} = -B^{-1} \frac{\partial B}{\partial x} B^{-1} \tag{14}$$

En différenciant alors L_t dans (12) par rapport à un paramètre θ_i et en utilisant (13) et (14), on obtient :

$$\frac{\partial L_t}{\partial \theta_i} = -\frac{1}{2} \operatorname{Tr}\left(H^{-1} \frac{\partial H_t}{\partial \theta_i}\right) - \left(\frac{\partial \eta_t}{\partial \theta_i}\right)' H_t^{-1} \eta_t + \frac{1}{2} \eta_t H^{-1} \frac{\partial H_t}{\partial \theta_i} H^{-1} \eta_t \tag{15}$$

expression que l'on peut réécrire en notant que le dernier terme est égal à sa trace et en utilisant la propriété de commutativité de la trace :

$$\frac{\partial L_t}{\partial \theta_i} = -\frac{1}{2} \operatorname{Tr}\left\{\left(H_t^{-1} \frac{\partial H_t}{\partial \theta_i}\right)(I - H_t^{-1} \eta_t \eta_t')\right\} - \left(\frac{\partial \eta_t}{\partial \theta_i}\right)' H_t^{-1} \eta_t$$

$$:= \qquad M_t \qquad + \qquad N_t$$

Pour obtenir la matrice des dérivées secondes de la log-vraisemblance, on peut écrire :

$$\frac{\partial M_t}{\partial \theta_j} = -\frac{1}{2} \operatorname{Tr}\left\{\left[\partial \left(H_t^{-1} \frac{\partial H_t}{\partial \theta_i}\right) \Big/ \partial \theta_j\right](I - H_t^{-1} \eta_t \eta_t')\right\}$$

$$-\frac{1}{2} \operatorname{Tr}\left(H_t^{-1} \frac{\partial H_t}{\partial \theta_i} H_t^{-1} \frac{\partial H_t}{\partial \theta_j} H_t^{-1} \eta_t \eta_t'\right)$$

$$+\frac{1}{2} \operatorname{Tr}\left\{H_t^{-1} \frac{\partial H_t}{\partial \theta_i} H_t^{-1} \left[\frac{\partial \eta_t}{\partial \theta_j} \eta_t' + \eta_t \left(\frac{\partial \eta_t}{\partial \theta_j}\right)'\right]\right\} \tag{16}$$

Comme les seules variables aléatoires de cette dernière expression sont les η_t, le premier terme s'annule et les deux dernières matrices du deuxième terme se neutralisent lorsqu'on prend l'espérance de (16). On voit d'après (8) que $\partial \eta_t / \partial \theta_i$ dépend seulement des innovations passées (par l'intermédiaire de $x_{t|t-1}$) et des variables exogènes à la période présente. Etant donné que, compte tenu de l'hypothèse de normalité, les innovations passées et présentes sont indépendantes, le troisième terme de l'expression (16) disparait lui aussi en espérance. On a donc au total :

$$E \frac{\partial M_t}{\partial \theta_j} = -\frac{1}{2} \operatorname{Tr}\left(H_t^{-1} \frac{\partial H_t}{\partial \theta_i} H_t^{-1} \frac{\partial H_t}{\partial \theta_j}\right) \tag{17}$$

En différenciant maintenant N_t par rapport à θ_j, on obtient :

$$\frac{\partial N_t}{\partial \theta_j} = -\frac{\partial^2 \eta_t}{\partial \theta_i \partial \theta_j} H_t^{-1} \eta_t - \left(\frac{\partial \eta_t}{\partial \theta_i}\right)' \frac{\partial H^{-1}}{\partial \theta_j} \eta_t - \left(\frac{\partial \eta_t}{\partial \theta_i}\right)' H^{-1} \frac{\partial \eta_t}{\partial \theta_j} \tag{18}$$

Les deux premiers termes disparaissent lorsqu'on passe en espérance. Le troisième dépend uniquement des innovations passées et est donc égal à son espérance conditionnée par les observations passées, ce qui donne :

$$E \frac{\partial N_t}{\partial \theta_i} = -\left(\frac{\partial \eta_t}{\partial \theta_i}\right)' H_t^{-1} \frac{\partial \eta_t}{\partial \theta_j} \tag{19}$$

Le ij-ème élément de la matrice d'information est l'opposé de la somme de (17) et (19), elle-même sommée sur toutes les périodes t :

$$\mathcal{I}_{ij} = \sum_t \frac{1}{2} \mathrm{Tr} \left(H^{-1} \frac{\partial H_t}{\partial \theta_i} H_t^{-1} \frac{\partial H_t}{\partial \theta_j} \right) + \sum_t \left(\frac{\partial \eta_t}{\partial \theta_i} \right)' H_t^{-1} \frac{\partial \eta_t}{\partial \theta_j} \qquad (20)$$

Lorsqu'il n'existe pas de distribution a priori non dégénérée pour $x_{0|0}$ et que le processus x_t n'est pas stationnaire, les valeurs de $x_{0|0}$ et $P_{0|0}$ peuvent être considérées comme des paramètres incidents qu'on estime en maximisant la vraisemblance. Heureusement, on peut concentrer la vraisemblance par rapport à ces estimateurs, si bien que les éléments de $x_{0|0}$ et $P_{0|0}$ n'apparaissent pas dans θ. On trouve dans Rosenberg [31] une méthode pour concentrer la vraisemblance par rapport à ces valeurs initiales. Une méthode équivalente pour les estimer consiste à utiliser un algorithme de point fixe lissé (cf. Anderson et Moore [1]). Cet algorithme de lissage est semblable à l'algorithme présenté ci-dessus en (6) - (11), et il fournit des estimations $x_{t|T}$ et $P_{t|T}$, c'est-à-dire des estimations fondées sur l'ensemble de l'échantillon observé. Pour concentrer la vraisemblance par rapport aux valeurs initiales, on utilise simplement les estimations lissées, l'algorithme de lissage étant amorcé par une distribution a priori non informative. Les estimations qui en résultent sont les estimations du maximum de vraisemblance des valeurs initiales fondées sur l'ensemble de l'échantillon observé et conditionnées par les valeurs des paramètres.

Cette façon de résoudre le problème des valeurs initiales apparaît plus séduisante que celle consistant à se donner une distribution a priori non informative pour les processus X non stationnaires. Ces différents estimateurs ne diffèrent certes que dans les échantillons de taille finie ; mais, comme c'est le cas dans Engle [15], il peut être parfois souhaitable d'obtenir exactement les estimateurs du maximum de vraisemblance sous des hypothèses particulières.

5. TESTS DE DIAGNOSTIC (OU DE SPÉCIFICATION)

On dispose de trois grandes catégories de diagnostics pour évaluer l'adéquation d'un modèle aux données. Tout d'abord, on peut déduire de la matrice d'information des écarts-types estimés pour les paramètres estimés du modèle, ce qui peut permettre de juger si l'on a introduit trop de paramètres ou de détecter d'autres raisons pour lesquelles le modèle est inadéquat. Ensuite, on peut construire un test de multiplicateur de Lagrange pour tester si l'on a introduit assez de paramètres ; on en verra un exemple ci-dessous.

Enfin, on peut utiliser les innovations pour tester si les perturbations constituent un bruit blanc et ne sont pas intercorrélés. On est amené, en effet, à conclure à une erreur dans la spécification du modèle si l'on obtient une autocorrelation significative pour une quelconque des innovations ou des intercorrelations entre innovations relatives à différentes variables endogènes et à différentes périodes d'observations. Si l'on observe une autocorrelation significative, alors le modèle doit être augmenté d'une façon ou d'une autre. Si cette autocorrelation ne concerne qu'une seule innovation, alors c'est vraisemblablement que les perturbations de l'équation de mesure correspondante sont autocorrélées. Si l'on observe des intercorrélations entre innovations à des dates différentes, c'est que les variables explicatives communes aux équations de mesure ne sont pas suffisamment bien spécifiées ; ou bien elles doivent être redéfinies et intervenir différemment dans les équations de mesure, ou peut-être convient-il d'introduire d'autres variables explicatives communes.

Le recours au test du multiplicateur de Lagrange (ou test des scores) se répand de plus en plus en économétrie : cf, par exemple, Godfrey [18], Pagan [30], Breusch et Pagan [2], Engle [10] et Byron [5]. On peut trouver une bonne vue d'ensemble sur le principe de ce test dans Engle [12]. Ce test est asymptotiquement équivalent au test de Wald et au test du rapport de vraisemblance, et il est souvent beaucoup plus simple à mettre en œuvre puisque l'estimation n'a à être effectuée que dans l'hypothèse nulle. On peut aisément construire un test des scores destiné à tester si l'on a introduit suffisamment de paramètres dans le modèle DYMIMIC.

Supposons que l'on soupçonne que le vecteur θ des paramètres doive être augmenté, et que la spécification correcte corresponde au vecteur $\omega = \begin{bmatrix} \theta \\ \pi \end{bmatrix}$ avec rang $\pi = r$. On pourrait par exemple soupçonner que les perturbations doivent suivre un processus vectoriel autorégressif d'ordre 1, auquel cas π serait constitué des paramètres de ce processus. L'hypothèse nulle et l'alternative peuvent alors s'écrire :

$$H_o \quad : \quad \pi = 0$$

$$H_a \quad : \quad \pi \neq 0$$

Soit $\hat{\theta}$ l'estimateur du maximum de vraisemblance de θ sous H_o. Si H_o est vérifiée, $\left.\frac{\partial L}{\partial \omega}\right|_{\substack{\theta=\theta \\ \pi=0}}$ doit être «proche» de zéro : c'est sur cette remarque qu'est fondé le test. On peut démontrer que, sous certaines conditions de régularité, la statistique

$$\xi := \left[\left(\frac{\partial L}{\partial \omega} \right)' \mathcal{I}^{-1} \frac{\partial L}{\partial \omega} \right]_{\substack{\theta=\hat{\theta} \\ \pi=0}} \qquad (21)$$

est distribuée asymptotiquement comme un χ^2 à r degrés de liberté dans l'hypothèse H_o. Etant donné que $\hat{\theta}$ est l'estimateur du maximum de vraisemblance de θ

$$\left.\frac{\partial L}{\partial \theta}\right|_{\substack{\theta=\hat{\theta} \\ \pi=0}} = 0$$

Aussi pour construire le test, on a seulement besoin de calculer $\left.\frac{\partial L}{\partial \pi}\right|_{\substack{\theta=\hat{\theta} \\ \pi=0}}$ et les éléments additionnels de $\mathcal{J}$. On peut calculer numériquement $\left.\frac{\partial L}{\partial \pi}\right|_{\substack{\theta=\hat{\theta} \\ \pi=0}}$ de façon simple en effectuant r passes à travers le filtre de Kalman, et les éléments additionnels de $\mathcal{J}$ peuvent être évalués en utilisant l'expression donnée dans la partie précédente. Arrivé à ce point, toutes les dérivées premières de η et de H sont disponibles sans calcul supplémentaire ; aussi les tests sont-ils très peu coûteux à effectuer.

6. APPLICATION ÉCONOMIQUE

Dans ce paragraphe, les techniques décrites ci-dessus vont être appliquées à l'estimation d'un facteur commun non observé pour des données relatives aux taux de salaires dans plusieurs secteurs d'activité d'une zone urbaine (pour plus de détails, se reporter à Engle et Watson [14], [15]). On suppose que le taux de salaire à Los Angelès dépend de facteurs nationaux spécifiques à chaque secteur d'activité, d'un facteur spécifique à Los Angelès et commun à tous les secteurs, et enfin de facteurs qui sont spécifiques à la fois à la région de Los Angelès et à chaque secteur. On se propose de construire une série de salaires pour Los Angelès et, à partir de là, d'examiner si ces salaires ont tendance à augmenter ou à diminuer par rapport à ceux relatifs aux Etats-Unis dans leur ensemble.

On a de bonnes raisons économiques pour penser que l'influence nationale comme l'influence locale jouent un rôle important. Dans un monde parfaitement néoclassique, tous les salaires sur un marché du travail donné évolueraient parallèlement pour équilibrer le marché ; l'influence nationale ne jouerait donc aucun rôle. D'un autre côté, les salaires dans un secteur d'activité d'extension nationale tiendraient compte des conditions économiques relatives à ce secteur en raison du capital humain qui le caractérise, des syndicats et des contrats nationaux. Et même sans ces contraintes institutionnelles, on pourrait s'attendre à ce que les secteurs qui sont confrontés

à une demande élastique par rapport aux prix et qui travaillent pour le marché national soient peu sensibles aux marchés locaux du travail : si un marché local du travail s'avérait étroit, un tel secteur réduirait vraisemblablement sa production locale pour la transférer dans une région où le marché local du travail lui serait plus favorable. Ainsi, la mobilité du capital et la fixité du prix de vente des produits rendraient les taux de salaire observés localement peu sensibles aux marchés locaux du travail. En bref, on pourrait s'attendre à ce que des secteurs d'activité à vocation nationale comme l'industrie et peut-être le commerce de gros soient peu sensibles aux conditions locales du marché du travail, tandis que le commerce de détail et le bâtiment le soient bien davantage (pour plus de détails et d'autres arguments, cf. Engle [9], [13]).

La spécification retenue pour le modèle économétrique est très simple. Soit w_{it} et n_{it} le logarithme du taux de salaire observé dans le secteur i l'année t pour Los Angelès et l'ensemble des Etats-Unis respectivement. Le logarithme du taux de salaire pour la région de Los Angelès est m_t, et les e_{it} sont des perturbations normales. Toutes les données sont centrées. Pour chacun des cinq secteurs retenus, le modèle s'écrit :

$$w_{it} = a_i m_t + b_i n_{it} + e_{it} \tag{22}$$

Ce sont là les équations de mesure qui correspondent aux équations (2). Les a_i sont les poids pour chaque secteur i du facteur local non observé «Los Angelès» ; ils devraient être faibles mais encore positifs pour les secteurs qui produisent pour le marché national. Le poids du facteur local dans la construction est pris égal à 1 à titre de normalisation.

Il faut aussi une équation pour expliquer m. On a considéré plusieurs spécifications alternatives. Dans la première, on suppose que m est fonction du taux de chômage local, et le modèle résultant a donc une structure DYMIMIC. Tout en donnant des estimations raisonnables des a_i et des b_i, ce modèle a systématiquement donné au taux de chômage un coefficient du mauvais signe, ce qui est sans doute dû au biais de simultanéité résultant de l'endogénéité du taux de chômage au niveau local.

Faute de pouvoir imaginer des variables instrumentales satisfaisantes, on a formulé un deuxième modèle, qualifié de modèle à effets fixes. Dans celui-ci, le facteur local m est simplement repéré par une variable indicatrice pour chaque année. Si ce modèle donne des résultats raisonnables et conduit à une évolution très régulière des $\hat{m}_t$, il fait usage d'un grand nombre de paramètres et est évidemment inutilisable pour la prévision.

Le troisième modèle permet à m_t de suivre un simple processus autoregressif du premier ordre. Ce n'est donc qu'un modèle d'analyse factorielle dynamique. Les tests de diagnostic présentés au paragraphe 5 ont révélé dans les innovations d'importantes autocorrélations et intercorrélations à des dates différentes, démontrant ainsi l'inadéquation du modèle. Aussi

celui-ci a-t-il été reformulé pour permettre à m_t de suivre un processus autorégressif du second ordre et à chacune des perturbations e_{it} $(i = 1, \ldots, 5)$ un processus autoregressif du premier ordre. En appelant e_t le vecteur (5, 1) des perturbations à la date t et a le vecteur (5, 1) des poids a_i du facteur local, ce modèle d'analyse factorielle dynamique peut être écrit sous forme d'un modèle dans l'espace des états de la façon suivante :

$$\begin{bmatrix} m_t \\ m_{t-1} \\ e_t \end{bmatrix} = \left[\begin{array}{cc:c:c} \varphi_1 & \varphi_2 & & 0 \\ 1 & 0 & & \\ \hdashline & & \rho_1 & \\ 0 & & & \ddots \\ \hdashline & & 0 & \\ & & & \rho_5 \end{array}\right] \begin{bmatrix} m_{t-1} \\ m_{t-2} \\ e_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ 0 \\ v_t \end{bmatrix}$$

$$w_t = [a, 0, I] \begin{bmatrix} m_t \\ m_{t-1} \\ e_t \end{bmatrix} + \begin{bmatrix} b_1 & & 0 \\ & \ddots & \\ 0 & & b_5 \end{bmatrix} n_t \qquad (23)$$

$$Q = \left[\begin{array}{cc:c:c} \sigma_\epsilon^2 & 0 & & 0 \\ 0 & 0 & & \\ \hdashline & & \sigma_1^2 & \\ 0 & & & \ddots \\ \hdashline & & 0 & \\ & & & \sigma_5^2 \end{array}\right] \qquad R = 0$$

Les résultats de l'estimation de ce modèle sont donnés dans le Tableau 1.

Les poids du facteur local les plus élevés se trouvent dans le bâtiment et dans le commerce de détail. Tous les poids obtenus sont positifs et, à en juger par leurs écarts-types estimés, ils sont tous très nettement significativement différents de zéro. Les écarts-types estimés de chaque équation sont au plus de l'ordre de 1 %. Les racines de l'équation caractéristique du processus autorégressif du second ordre estimé pour le facteur local étant approximativement 1 et 0,6, la possibilité d'une racine unitaire indique simplement que, sur la période d'observation, il n'existait sans doute pas de mécanisme rééquilibrant susceptible d'amener le taux de salaire de Los Angelès à égalité avec celui prévalant dans le reste du pays. (Lorsqu'on trouve une racine unitaire dans un tel modèle, il n'y a pas de changement dans les procédures d'estimation et d'inférence à ceci près que les conditions initiales doivent être considérées comme des paramètres incidents et elles-mêmes estimées). Des diagnostics de confirmation fondés sur les autocorrélations et les intercorrélations des innovations à des dates différentes ont été effectués à partir de ces derniers résultats, et on a trouvé que l'autocorrélation avait bien été éliminée. Des prévisions hors-échantillon ont été effectuées à titre de comparaison avec des régressions dynamiques pour chaque secteur séparément : en termes de fonction objectif jointe, elles

Tableau 1

Analyse factorielle dynamique

$m_t = \varphi_1 m_{t-1} + \varphi_2 m_{t-2} + e_t$

$w_{it} = a_i m_t + b_i n_{it} + e_{it}$ $\quad i = 1, \ldots, 5$ (secteurs)

$e_{it} = \rho_i e_{it-1} + v_{it}$

(Ecarts-types estimés entre parenthèses)

	$\hat{a}_i$	$\hat{b}_i$	$\hat{\rho}_i$	$\hat{\sigma}_i^2 \times 10^4$	$\hat{\sigma}_i$
i=1: Bâtiment et Travaux publics	1	0,874 (0,078)	0,628 (0,389)	0,598 (0,329)	0,008
i=2: Biens durables	0,549 (0,090)	0,786 (0,053)	0,742 (0,155)	0,835 (0,266)	0,009
i=3: Biens non durables	0,380 (0,091)	0,786 (0,040)	0,898 (0,107)	0,466 (0,149)	0,007
i=4: Commerce de gros	0,302 (0,075)	0,959 (0,032)	0,519 (0,227)	1,191 (0,352)	0,011
i=5: Commerce de détail	0,663 (0,070)	0,810 (0,059)	0,340 (0,289)	0,941 (0,343)	0,010

	$\hat{\varphi}_1$	$\hat{\varphi}_2$	$\hat{\sigma}_\epsilon^2 \times 10^4$	$\hat{\sigma}_\epsilon$
Composante locale	1,606 (0,125)	− 0,619 (0,145)	1,229 (0,585)	0,011

ont montré systématiquement la supériorité des prévisions issues du modèle d'analyse factorielle dynamique. Les estimations de m_t révèlent une croissance régulière de ce facteur local, sauf à la fin des années 1960 et au début des années 1970 où se manifeste un ralentissement qui pourrait être attribué à l'achèvement des contrats aérospatiaux et militaires.

Pour conclure, le modèle retenu apparaît satisfaisant à la fois du point de vue économique et du point de vue statistique. Les méthodes d'estimation et de diagnostic permettent effectivement de choisir entre différentes spécifications, et l'utilisation du filtre de Kalman et de la méthode des scores se révèle tout-à-fait praticable avec un problème de cette taille.

REFERENCES BIBLIOGRAPHIQUES

[1] ANDERSON B.D.O. and MOORE J.B. – *Optimal Filtering*, New Jersey : Prentice Hall, 1979.

[2] BREUSCH T.S. and PAGAN A.R. – «The Lagrange Multiplier Test and Its Applications to Model Specification in Econometrics», *Review of Economic Studies*, 47, 239-253, 1980.

[3] BRILLINGER D.R. – *Time Series Data Analysis and Theory*, New York : Holt, Rinehart and Winston, Inc, 1975.

[4] BROWN R.L. DURBIN J., and EVANS J.M. – «Techniques for Testing the Constancy of Regression Relationships Over Time, with Comments», *Journal of the Royal Statistical Society*, Ser. B. 37, 149-192, 1975.

[5] BYRON R.P. – «The Restricted Aitken Estimation of Sets of Demand Relations», *Econometrica*, 38, 816-830, 1970.

[6] CHOW G.C. – *Analysis and Control of Dynamic Economic Systems*, New York : John Wiley and Sons, inc, 1975.

[7] COOLEY T.F. and PRESCOTT E.C. – «Tests of an Adaptive Regression Model», *Review of Economics and Statistics*, 55, 248-256, 1973.

[8] COOLEY T.F. and PRESCOTT E.C. – «Estimation in the Presence of Stochatic Parameter Variations», *Econometrica*, 44, 167-184, 1976.

[9] ENGLE R.F. – «Issues in the Specification of an Econometric Model of Metropolitan Growth», *Journal of Urban Economics*, 1, 250-267, 1974.

[10] ENGLE R.F. – «Hypothesis Testing in Spectral Regression : The Lagrange Multiplier as a Regression Diagnostic», in *Criteria for Evaluation of Econometric Models*, ed. J. Kmenta and J. Ramsey, NBER Conference Proceedings, Ann Arbor, 1977.

[11] ENGLE R.F. – «Estimating Structural Models of Seasonality», in *Seasonal Analysis of Economic Time Series*, Economic Research Report, ER-1, U.S. Dept. of Commerce, Bureau of the Census, 1979.

[12] ENGLE R.F. – «A General Approach to the Construction of Model Diagnostics Based Upon the Lagrange Multiplier Principle», DP # 156, Social Science Research Council, 1979 a.

[13] ENGLE R.F. – «Estimation of the Price Elasticity of Demand Facing Metropolitan Producers», *Journal of Urban Economics*, 6, 42-64, 1979 b).

[14] ENGLE R.F. and WATSON M.W. – «A One Factor Multivariate Time Series Model of Metropolitan Wage Rates», Discussion Paper 78-15, Department of Economics, University of California, San Diego, 1978.

[15] ENGLE R.F. and WATSON M.W. – «A One Factor Multivariate Time Series of Metropolitan Wage Rates», revised, forthcoming in the *Journal of the American Statistical Association*, 1979.

[16] GARBADE K. – «Two Methods for Examining the Stability of Regression Coefficients», *Journal of the American Statistical Association*, 72, 54-63, 1977.

[17] GEWEKE J. – «The Dynamic Factor Analysis of Economic Time-Series Models», in *Latent Variables in Socio-Economic Models*, eds. D.J. Aigner and A.S. Goldberger, New York : North-Holland, 1977.

[18] GODFREY L.G. – «Testing Against General Autoregressive and Moving Average Error Models when the Regressors Include Lagged Dependent Variables», *Econometrica*, 46, 1293-1301, 1978.

[19] GOLDBERGER A.S. – «Structural Equation Methods in the Social Sciences», *Econometrica*, 40, 979-1001, 1972.

[20] GOLDBERGER A.S. – «Maximum-likelihood Estimation of Regressions Containing Unobservable Independent Variables», in *Latent Variables in Socio-Economic Models*, eds D.J. Aigner and A.S. Goldberger, New York : North-Holland, 1977.

[21] GRANGER C.W.J. and NEWBOLD P. – *Forecasting Economic Time Series*, New York : Academic Press, 1977.

[22] GUPTA N.K. and MEHRA R.K. – «Computational Aspects of Maximum Likelihood Estimation and Reduction in Sensitivity Function Calculation», *I.E.E.E. Transactions on Automatic Control*, AC-19, 774-783, 1974.

[23] HANNAN E.J. – «The Identification and Parameterization of Armax and State Space Forms», *Econometrica*, 44, 713-724, 1976.

[24] HARVEY A. and COLLIER P. – «Testing for Functional Misspecification in Regression Analysis», *Journal of Econometrics*, 6, 103-120, 1977.

[25] HARVEY A. and PHILLIPS G.D.A. – «The Estimation of Regression Models with Autoregressive-Moving Average Disturbances», *Biometrika*, 66, 49-58, 1979.

[26] MEHRA R.K. – «Identification in Control an Economics : Similarities and Differences», *Annals of Economic and Social Measurement*, 3, 21-47, 1974.

[27] MORRISON D.F. – *Multivariate Statistic Methods*, New York : McGraw Hill, 1976.

[28] NERLOVE M. – «Analysis of Economic Time-Series by Box-Jenkins and Related Techniques», Report 7156, Center for Mathematical Studies in Business and Economics, University of Chicago, 1971.

[29] PAGAN A. – «A Note on the Extraction of Components from time Series», *Econometrica*, 43, 163-168, 1975.

[30] PAGAN A. – «A Unified Approach to Estimation and inference for Stochastically Varying Coefficient Regression Models», Discussion Paper 7814, Center for Operations Research and Econometrics, 1978.

[31] ROSENBERG B. – «The Analysis of a Cross Section of Time Series by Stochastically Convergent Parameter Regression», *Annals of Economic and Social Measurement*, 2, 299-428, 1973.

[32] SARGENT T.J. and SIMS C.A. – «Business Cycle Modeling without Pretending to have too much *a priori* Economic Theory», in *New Methods in Business Cycle Research : Proceedings from a Conference*, Federal Reserve Bank of Minneapolis, 1977.

[33] SIMS C.A. – «Macro-economics and Reality», Discussion Paper No. 77-91, Center for Economic Research, University of Minnesota, 1977.

[34] TAYLOR L. – «The Existence of Optimal Distributed Lags», *Review of Economic Studies*, 37, 95-106, 1970.

[35] WALLIS K.F. – «Multiple Time Series Analysis and the Final Form of Econometric Models», *Econometrica*, 45, 1481-1498, 1977.

[36] ZELLNER A. – «Estimation of Regression Relationships Containing Unobservable Variables», *International Economic Review*, 11, 441-454, 1970.