

LES MODÈLES À RETARDS ÉCHELONNÉS :
FONDEMENTS THÉORIQUES, SPECIFICATIONS
ET MÉTHODES D'ESTIMATIONS USUELLES*

Par

Pierre-Alain MUET

N°7901

* Cette étude a été préparée pour le compte de la Direction de la Prévision.

SOMMAIRE

	<u>Pages</u>
1 - LES FONDEMENTS THEORIQUES	2
1.1. La formation des anticipations	2
1.2. Anticipations et prévision optimale	3
1.3. Grandeur permanente et transitoire : la distribu- tion géométrique	5
1.4. Effet de stock ou d'habitude : le modèle dyna- mique de demande	8
1.5. Délais d'ajustement	8
1.6. Fonction globale des retards	11
2 - UNE FORMULATION GENERALE DES MODELES A RETARDS ECHELONNES	12
2.1. Définition et propriétés des opérateurs de déca- lage	12
2.2. Fonction génératrice, délai moyen	13
2.3. Distributions finies, infinies et formes autoregres- sives	14
2.4. Temps discret, temps continu	16
3 - LES DISTRIBUTIONS INFINIES ET LEUR ESTIMATION SOUS FORME AUTOREGRESSIVE (I)	17
3.1. La distribution géométrique estimée sous forme autoregressive : quelques exemples	21
3.2. Distributions du premier ordre : l'accélérateur flexible sur données annuelles	27
3.3. Distributions du second ordre	30
3.4. Comparaison des distributions du premier et du second ordre : accélérateur flexible sur série chronologique annuelle (1953-1974).	32
3.5. Modèles autoregressifs à plusieurs variables....	36
4 - LES DISTRIBUTIONS INFINIES ET LEUR ESTIMATION SOUS FORME AUTOREGRESSIVE (II) : PROBLEMES D'IDENTIFI- CATION ET D'ESTIMATION	41
4.1. Spécification incomplète ou erronée du modèle ..	42
4.2. Les biais introduits par l'estimation sur petits échantillons	44
4.3. Corrélacion temporelle des erreurs	46
4.4. Méthodes d'estimation alternatives	48
4.5. Un exemple : distribution du second ordre de l'accélérateur flexible.....	55

	<u>Pages</u>
5 - LES DISTRIBUTIONS FINIES ET LEUR APPROXIMATION POLYNOMIALE	56
5.1. La distribution polynômiale et son estimation (méthode d'Almon)	57
5.2. Un exemple : accélérateur flexible sur données trimestrielles.....	59
5.3. Méthode d'ALMON ou moindres carrés : un test de l'hypothèse polynômiale	62
5.4. Limites de la méthode d'ALMON	64
5.5. Compléments : distributions à coefficients variables ou stochastiques	67
CONCLUSION	68
REFERENCES BIBLIOGRAPHIQUES	71

L'étude des modèles à retards échelonnés s'est considérablement développée au cours des vingt dernières années. Depuis les premiers travaux de KOYCK [1954], JORGENSON [1966] et ALMON [1965], de nombreuses méthodes d'estimations ont été développées sans toutefois correspondre nécessairement aux besoins de l'économiste ou du "modélisateur". En outre, depuis le remarquable "survey" de GRILICHES [1967], peu d'articles ou d'ouvrages ont été consacrés à une synthèse opérationnelle des divers modèles à retards échelonnés*. Une telle synthèse s'avère délicate en raison de la coupure croissante entre le développement de la "théorie économétrique" -c'est-à-dire les méthodes statistiques de l'économétrie-, et l'application de ces mêmes méthodes à la quantification de phénomènes économiques (appelée parfois économétrie "empirique"). La présente étude ne prétend nullement combler ce fossé mais se propose, plus modestement, de traiter des méthodes d'estimations usuelles des modèles à retards échelonnés. Ce parti-pris peut nous avoir conduit à passer sous silence des méthodes qui ont connu encore peu d'applications, mais présentent un intérêt certain pour les études empiriques. Aussi accueillerons-nous avec beaucoup d'intérêt les suggestions et compléments d'informations qui pourraient nous être apportés.

Notre étude comprend cinq parties.

Les fondements théoriques des modèles à retards échelonnés sont rappelés dans la première partie. La seconde présente la spécification générale de ces modèles, en utilisant notamment la symbolique et les propriétés des opérateurs de décalage. Enfin les trois dernières parties sont consacrées à l'étude des méthodes d'estimations économétriques. Toutes les méthodes présentées font l'objet d'une application utilisant les séries annuelles ou trimestrielles de la comptabilité nationale française.

./.

* L'ouvrage de DHRYMES [1971.] apporte des démonstrations intéressantes mais est de peu d'usage pour l'économiste car il ne différencie pas les méthodes opérationnelles de celles qui ne présentent strictement aucun intérêt empirique.

L'estimation sous forme autoregressive des distributions infinies (géométrique, premier et deuxième ordre...) est étudiée dans les troisième et quatrième partie, tandis que la dernière partie examine les applications et les limites des distributions polynômiales.

1 - LES FONDEMENTS THEORIQUES

Les retards échelonnés des relations économétriques ont deux justifications principales : ils représentent soit la formation des anticipations à partir des observations passées des grandeurs sur lesquelles s'appuie la décision des agents économiques, soit au contraire des délais concrets de réalisation de ces décisions.

1.1. La formation des anticipations

Le processus de formation des anticipations peut être rattaché au problème de la prévision d'une grandeur future à partir de diverses informations qui peuvent être soit l'évolution passée et prévue d'un certain nombre de variables qui influencent la grandeur concernée, soit de façon plus restrictive les seules observations passées de la grandeur en question. Dans ce dernier cas l'anticipation réalisée à la date t pour la période $(t+\theta)$ s'exprimera généralement sous la forme d'une fonction à retards échelonnés de ces observations passées :

$$\hat{x}_t^{t+\theta} = f(x_{t-1}, x_{t-2}, x_{t-3} \dots)$$

Une première formulation des anticipations, appuyée sur l'hypothèse de leur adaptation progressive aux réalisations, a été proposée par GAGAN [1956]. Ce dernier a supposé que l'agent économique révisait ses anticipations (x_t^*) à chaque période, en fonction de l'écart précédemment observé entre la grandeur anticipée et réalisée $(x_{t-1}^* - x_{t-1})$. Si l'on suppose par exemple que l'on reporte dans la nouvelle prévision x_t^* une fraction constante (η) de l'erreur précédemment réalisée :

$$(x_t^* - x_{t-1}^*) = \eta (x_{t-1}^* - x_{t-1})$$

On obtient une anticipation qui s'exprime sous la forme d'une distribution géométrique de retards des grandeurs passées :

$$x_t^* = \sum_{i=1}^{\infty} \eta (1-\eta)^{i-1} x_{t-i}$$

Une justification plus rigoureuse de la distribution géométrique a été développée par MUTH pour représenter la notion de grandeur permanente que l'on trouve aussi bien pour justifier les retards échelonnés des fonctions de consommation que l'investissement.

1.2. Anticipations et prévision optimale

Une façon simple de représenter le processus de formation des anticipations consiste à l'assimiler à la prévision optimale des réalisations futures de la grandeur x_t , à partir des observations passées (x_{t-i}) de cette grandeur. En général l'évolution des grandeurs économiques présente une certaine régularité comme par exemple un caractère plus ou moins cyclique. On peut donc les assimiler à des processus stationnaires. Ces processus peuvent être eux-mêmes décomposés en un processus déterminable (x_t^d) et un processus indéterminable (ξ_t) que nous supposerons de moyenne nulle en longue période. Si ce processus indéterminable est purement aléatoire, et notamment sans autocorrélation, la prévision optimale* pour la période $(t+\theta)$ s'identifiera à la valeur de la composante déterminable du processus :

$$\hat{x}_t^{t+\theta} = x_{t+\theta}^d$$

Si par exemple la grandeur x_t peut être représentée par un trend temporel $\varphi(t)$ augmenté de variations (ξ_t) aléatoires autour du trend, la prévision optimale pour la période $(t+\theta)$ sera $\varphi(t+\theta)$. En général, la composante indéterminable présente une certaine autocorrélation, c'est-à-dire une certaine stationnarité. La prévision optimale pour la période $(t+\theta)$ ne s'identifie plus alors à la composante déterminable du processus car l'espérance mathématique de $\xi_{t+\theta}$ n'est pas nulle.

* La prévision $\hat{x}_t^{t+\theta}$ est optimale si elle minimise la somme des carrés des erreurs.

Comme la composante déterminable ne pose pas de problème de prévision, nous supposons par la suite que x_t s'identifie à sa composante indéterminable ξ_t . On peut montrer* que le processus indéterminable admet en général une représentation autoregressive du type :

$$x_t = \varepsilon_t + \sum_{j=1}^r b_j x_{t-j},$$

où ε_t est purement aléatoire .

A partir de n observations** passées du processus, on peut construire à la date t , des prévisions optimales pour les différentes périodes futures : \hat{x}_t^{t+1} , \hat{x}_t^{t+2} ... $\hat{x}_t^{t+\theta}$.

La prévision optimale pour $(t+1)$ est :

$$(1) \quad \hat{x}_t^{t+1} = \sum_{j=1}^n b_j x_{t-j},$$

et pour $t+\theta$:

$$(2) \quad \hat{x}_t^{t+\theta} = \sum_{i=1}^{\theta-1} b_i \hat{x}_t^{t+\theta-i} + \sum_{i=0}^n b_i x_{t+\theta-i}$$

En tenant compte de la récurrence, on peut exprimer la prévision $\hat{x}_t^{t+\theta}$ en fonction des seules valeurs passées $x_t, x_{t-1}, \dots, x_{t-i}$. On vérifie qu'en règle générale, les anticipations dépendent de la période de prévision θ . On obtient par exemple pour $\theta=2$:

$$(3) \quad \hat{x}_t^{t+2} = \sum_{i=1}^{\infty} (b_1 b_i + b_{i+1}) x_{t+1-i}$$

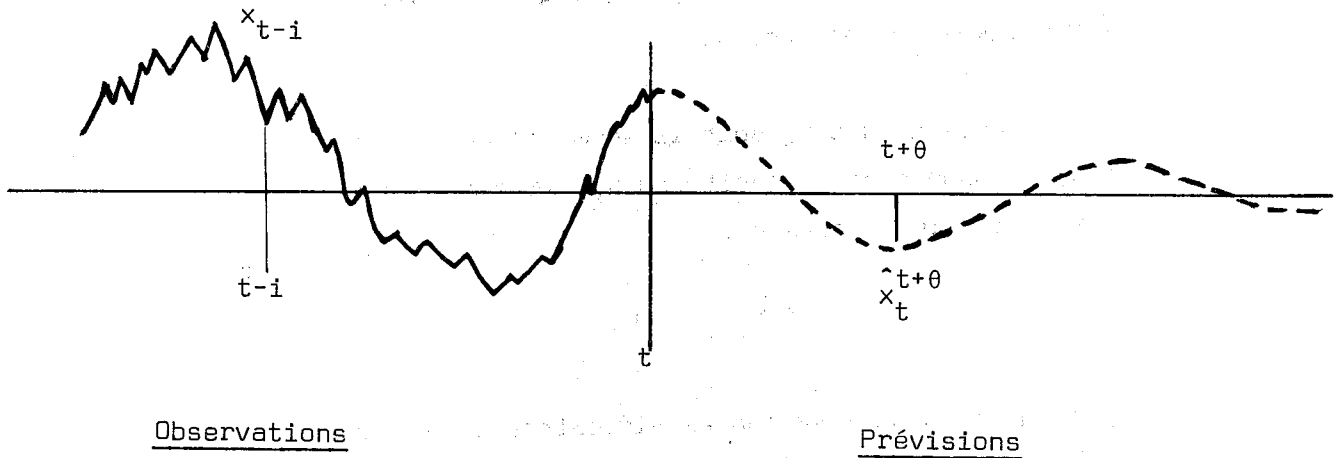
On notera cependant que quelle que soit la forme de la structure de retards b_j , les prévisions en longue période ($\theta \rightarrow \infty$) tendent vers la composante déterminable du processus (la prévision du processus indéterminable tend vers zéro), comme le montre le graphique (1).

./.

* Théorème de WOLD - voir par exemple MALINVAUD [1969] Chapitre 11.

** Lorsque n est infini, on peut se contenter en pratique d'un nombre fini d'observations car $b_j \rightarrow 0$ si $j \rightarrow \infty$.

GRAPHIQUE (1) - PREVISIONS OPTIMALES D'UN PROCESSUS STATIONNAIRE



1.3. Grandeur permanente et transitoire : la distribution géométrique

Nous venons de voir que les prévisions $\hat{x}_t^{t+\theta}$ dépendaient en général de la longueur θ de la période de prévision. On peut chercher, à quelles conditions, ces prévisions deviennent indépendantes de la période de prévision θ . On doit avoir en particulier :

$$\hat{x}_t^{t+1} = \hat{x}_t^{t+2}$$

d'où l'on déduit en comparant (1) et (3) :

$$(4) \quad b_i = (b_1 b_i + b_{i+1}),$$

et

$$b_{i+1} = (1-b_1) b_i = (1-b_1)^i b_1.$$

La distribution b_i est ainsi une distribution géométrique.

Elle correspond à la prévision optimale d'un processus stationnaire du type

$$(5) \quad x_t = \varepsilon_t + \lambda \sum_{i=1}^{\infty} \varepsilon_{t-i}$$

le processus comporte une composante transitoire ϵ_t et une composante permanente $\sum_{i=1}^{\infty} \epsilon_{t-i}$, qui a cependant la particularité d'être corrélée avec la réalisations passées de la variable transitoire. Si nous supposons au contraire les deux processus indépendants, on peut montrer que la distribution géométrique constitue encore une prévision optimale du processus.

Supposons que x_t soit la somme d'une composante permanente \bar{x}_t et d'une composante transitoire η_t , la composante permanente étant définie par le processus :

$$\bar{x}_t = \bar{x}_{t-1} + \epsilon_t$$

ϵ_t et η_t sont deux variables aléatoires de moyenne nulle et d'écart-type σ_ϵ et σ_η .

On démontre (MUTH (1960)) que la structure optimale des b_i est une distribution géométrique de retards :

$$b_i = (1 - \lambda) \lambda^i,$$

où le coefficient λ dépend de la variabilité respective des composantes permanentes et transitoires :

$$(6) \quad \lambda = 1 + \frac{1}{2} \frac{\sigma_\epsilon^2}{\sigma_\eta^2} - \frac{\sigma_\epsilon}{\sigma_\eta} \sqrt{1 + \frac{1}{4} \frac{\sigma_\epsilon^2}{\sigma_\eta^2}}$$

La relation (6) montre que si les changements de la composante permanente sont faibles par rapport à ceux de la composante transitoire, λ sera proche de un et les pondérations des observations passées seront presque identiques. Si au contraire les changements de la composante permanente sont importants, λ sera proche de zéro et les pondérations des périodes récentes seront beaucoup plus importantes.

La relation (6) resterait applicable si les processus η_t et ϵ_t étaient corrélés, il suffirait de changer $\frac{\sigma_\epsilon^2}{\sigma_\eta^2}$ par $\frac{\sigma_\epsilon^2}{\sigma_\eta^2 + \sigma_{\epsilon\eta}}$.

La distribution optimale étant géométrique, la prévision sera la même pour toutes les périodes futures $(t+1)$, ..., $(t+\theta)$, et égale à la composante permanente de la variable x_t .

$$\hat{x}_t^{t+\theta} = \dots = \hat{x}_t^{t+1} = \bar{x}_t$$

Lorsque la grandeur x_t croît en longue période à taux constant g , la grandeur permanente doit être actualisée par le taux de croissance g , puisque la prévision

$$\frac{\hat{x}_t^{t+\theta}}{(1+g)^\theta}$$
 est constante en longue période. On peut également

substituer à la notion de grandeur permanente x_t , la notion de croissance permanente, et appliquer les relations précédentes aux taux de croissance $\overset{0}{x}_t$.

On notera en conclusion que la distribution géométrique ne constitue une prévision optimale que lorsque le processus comporte une composante permanente au sens précédemment défini. Cependant, même lorsqu'elle n'est pas optimale (cas d'un processus stationnaire quelconque), la distribution géométrique reste une prévision relativement efficace comme l'a montré COX [1961].

Nous retiendrons plus généralement que l'on peut, en assimilant le processus d'anticipation à la prévision optimale d'une grandeur à partir des observations passées, représenter celui-ci par un modèle à retards échelonnés des observations passées :

$$\hat{x}_t^{t+\theta} = \sum_i a_{t-i}^\theta x_{t-i}$$

où les coefficients a_{t-i}^θ tendent généralement vers zéro lorsque i ou θ tend vers l'infini.

1.4. Effets de stock ou d'habitude : le modèle dynamique de demande

Une notion voisine de la grandeur permanente a été utilisée par HOUTHAKKER et TAYLOR [1968] et [1970] pour l'étude économétrique des fonctions de consommation.

HOUTHAKKER et TAYLOR considèrent que la consommation des ménages (C_t) résulte d'une part des variables observées au cours de la période (x_t), d'autre part des comportements antérieurs qu'ils résument par une variable d'état S_t :

$$(1) \quad C_t = a x_t + b S_t + d$$

Cette variable d'état représente, pour les biens durables, le stock disponible en début de période, et son influence est négative sur la consommation ($b < 0$). Pour les biens non durables, S_t traduit un effet d'habitude acquise avec l'ensemble des consommations passées et son influence est positive sur la consommation ($b > 0$). Dans les deux cas, la variable d'état peut s'exprimer en fonction des consommations antérieures :

$$(2) \quad S_t = (1-\delta) S_{t-1} + C_{t-1}$$

Dans le cas d'un stock δ est le taux de dépréciation du stock ; pour un effet d'habitude, δ représente un "taux d'oubli". En combinant les deux relations (1) et (2), on obtient le modèle dynamique de demande :

$$(3) \quad C_t = (1+b-\delta) C_{t-1} + a x_t - a (1-\delta) x_{t-1} + d \delta$$

La consommation dépend ainsi de la variable x_t (Revenu par exemple) selon une distribution du premier ordre. Ce modèle généralise le modèle d'ajustement au revenu permanent (distribution géométrique) que l'on obtiendrait en supposant $\delta=1$.

1.5. Délais d'ajustement

Nous avons vu que la première justification des retards échelonnés résultait de la formation des anticipations sur lesquelles s'appuient

les décisions des agents économiques. Le temps nécessaire à la réalisation des décisions constitue la seconde justification des modèles à retards échelonnés. Ce temps peut être subi par l'agent économique, par exemple le délai de livraison d'un bien commandé, ou résulter au contraire d'une décision délibérée. Ainsi en est-il des délais d'ajustement du capital et du travail justifiés par la recherche d'une trajectoire optimale d'ajustement aux grandeurs désirées, lorsque la variation des stocks capital et travail présente un coût croissant avec la vitesse d'ajustement. Développés initialement par EISNER et STROZ [1963] puis GOULD [1968] les modèles d'ajustement ont été étendus à plusieurs facteurs par LUCAS [1967], SCHRAMM [1970], TREADWAY [1971] MORTENSEN [1970] et les fonctions d'ajustement croisées ainsi obtenues ont été estimées par NADIRI et ROSEN [1974] et SCHRAMM [1970]. Pour éviter les complications inessentiels, nous aurons recours à un modèle très simplifié qui garde néanmoins l'idée de base de tous ces modèles d'ajustement, même si la formalisation que nous retenons n'est pas très rigoureuse. Supposons que l'emploi optimal anticipé pour la période t (N_t^*) soit supérieur à l'emploi de la période $t-1$. Si le changement du niveau de l'emploi de N_{t-1} à N_t^* ne comporte aucun coût, l'entreprise a intérêt à fixer son emploi, pour la période t , à la valeur optimale N_t^* (la valeur optimale est celle qui maximise son profit anticipé). Si elle ne s'ajuste pas à cette valeur N_t^* elle subira un manque à gagner qui est, en première approximation (1) une fonction quadratique de $(N_t - N_t^*)$:

$$(1) \quad q_1 = + c (N_t - N_t^*)^2$$

Supposons maintenant que l'ajustement comporte un coût proportionnel au carré de la vitesse d'ajustement $(N_t - N_{t-1})$:

$$(2) \quad q_2 = a (N_t - N_{t-1})^2$$

le coût total correspondant à une valeur N_t quelconque est égal à :

$$(3) \quad q(N_t) = \underbrace{c (N_t - N_t^*)^2}_{\text{Coût d'écart à l'optimum}} + \underbrace{a (N_t - N_{t-1})^2}_{\text{Coût d'ajustement}}$$

./.

(1) Si π_t^* désigne le profit maximum et $\varphi(N)$ la relation entre π et N on a, au voisinage de N_t^* : $(\pi_t - \pi_t^*) = \varphi'(N_t^*) \cdot (N_t - N_t^*) + \frac{1}{2} \varphi''(N_t^*) (N_t - N_t^*)^2 + \dots$
 Comme N_t^* est un maximum $\varphi' = 0$ $\varphi'' = -c \leq 0$, et la perte de profit est $c (N_t - N_t^*)^2$ en première approximation.

La minimisation de ce coût détermine l'emploi de la période t (N_t) :

$$\frac{d q (N_t)}{d N_t} = 2 c (N_t - N_t^*) + 2 a (N_t - N_{t-1}) = 0$$

L'ajustement optimal, c'est-à-dire l'ajustement qui minimise ce coût total sera une distribution géométrique :

$$(4) \quad N_t = \lambda N_{t-1} + (1-\lambda) N_t^* \quad \text{avec } \lambda = \frac{a}{c+a}$$

L'ajustement étant intertemporel, il faudrait en toute rigueur minimiser le coût actualisé ce qui ferait dépendre l'ajustement du taux d'intérêt (r). Ce modèle très simple montre néanmoins un résultat que l'on peut établir de façon plus générale: l'ajustement est d'autant plus lent que le coût d'ajustement (a) est élevé par rapport au coût de désajustement (c). Le délai moyen de l'ajustement est en effet :

$$\theta = \frac{\lambda}{1-\lambda} = \frac{a}{c}$$

Lorsque la fonction de production comporte plusieurs facteurs par exemple n facteur x_t^i et que la fonction de coût d'ajustement est également une fonction quadratique des vitesses d'ajustement des facteurs ($x_t^i - x_{t-1}^i$), le coût total est la forme quadratique :

$$(3') \quad q (X_t) = (X_t - X_t^*)' C (X_t - X_t^*) + (X_t - X_{t-1})' A (X_t - X_{t-1})$$

où C et A sont des matrices carrées symétriques.

La minimisation du coût total conduit à une équation d'ajustement vectorielle qui généralise la relation (4) :

$$X_t = \Lambda X_{t-1} + (I - \Lambda) X_t^*$$

avec $\Lambda = [C + A]^{-1} A$

Les fonctions d'ajustement font alors intervenir les "retards croisés" des différents facteurs, par exemple pour le capital et le travail :

$$\begin{bmatrix} K_t \\ N_t \end{bmatrix} = \begin{bmatrix} \lambda_1^1 & \lambda_1^2 \\ \lambda_2^1 & \lambda_2^2 \end{bmatrix} \begin{bmatrix} K_{t-1} \\ N_{t-1} \end{bmatrix} + \begin{bmatrix} 1-\lambda_1^1 & -\lambda_1^2 \\ -\lambda_2^1 & 1-\lambda_2^2 \end{bmatrix} \begin{bmatrix} K_t^* \\ N_t^* \end{bmatrix}$$

1.6. Fonction globale des retards

La détermination complète des retards échelonnés devrait donc faire intervenir à la fois les retards échelonnés des fonctions d'anticipation et les délais d'ajustement. Si ces deux distributions de retards étaient indépendantes, on obtiendrait la fonction globale en appliquant les règles de composition classique des retards (cf. paragraphe 2). En réalité anticipations et délais d'ajustement sont liés : en l'absence de coûts d'ajustement et dans le cas de facteurs substituables les décisions d'investissement et d'emploi ne nécessitent pas d'horizons prévisionnels éloignés puisque l'entreprise peut s'adapter instantanément aux grandeurs optimales. Si en revanche il existe des coûts d'ajustement et si a fortiori les techniques de production sont fixées sur les équipements installés (par exemple fonction de production putty-clay), les anticipations doivent concerner une période de temps relativement longue.

2 - UNE FORMULATION GENERALE DES MODELES A RETARDS ECHELONNES*

L'analyse formelle des modèles à retards échelonnés est grandement facilitée par le recours aux opérateurs de décalage.

La symbolique des opérateurs permet en effet de calculer simplement la résultante de la composition de plusieurs distributions de retards, ou les retards échelonnés correspondant à une formulation autoregressive.

2.1. Définition et propriétés des opérateurs de décalage

Définissons l'opérateur L par la relation

$$L x_t = x_{t-1}$$

Un retard d'ordre n résultera de n applications successives de l'opérateur soit :

$$L^n x_t = x_{t-n}$$

On voit sans difficultés que l'opérateur L possède les propriétés suivantes :

$$L (a x_t) = a L x_t$$

$$L (x_t + y_t) = L x_t + L y_t$$

Ces propriétés permettent d'écrire facilement un modèle à retards échelonnés sous la forme du produit d'un polynôme de l'opérateur L par la variable x_t . Le modèle :

$$y_t = a_0 x_t + a_1 x_{t-1} + \dots + a_n x_{t-n}$$

s'écrit ainsi :

$$y_t = \Phi(L) x_t$$

$$\text{avec } \Phi(L) = a_0 + a_1 L + a_2 L^2 + \dots + a_n L^n$$

./.

* Notre présentation s'inspire très largement de GRILICHES [1967].

Cette transformation des modèles à retards échelonnés permet de calculer facilement la structure des retards résultant de la composition de deux modèles à retards échelonnés. Supposons en effet que la variable y soit une fonction à retards échelonnés de x et que z soit une fonction à retards échelonnés de y :

$$y_t = \phi_1(L) x_t$$

$$z_t = \phi_2(L) y_t$$

Les propriétés de l'opérateur L (associativité, distributivité) permettent alors d'écrire :

$$z_t = \phi_2(L) [\phi_1(L) x_t] = [\phi_1(L) \cdot \phi_2(L)] x_t$$

Le polynôme représentant les retards échelonnés de la fonction globale s'obtient en effectuant le produit des polynômes de chacune des fonctions composantes. On remarquera que la somme des coefficients d'une structure de retards peut être calculée facilement à partir du polynôme $\phi(L)$. Cette somme qui représente le coefficient de réaction à long terme est en effet égale à $\phi(1)$:

$$S = a_0 + a_1 + \dots + a_n = \phi(1)$$

Nous allons poursuivre l'étude des propriétés de la fonction $\phi(L)$ en tirant parti de l'analogie formelle entre une distribution de retards et une distribution de probabilité.

2.2. Fonction génératrice, délai moyen

Considérons une distribution de retards échelonnés à coefficients positifs que nous normaliserons en divisant les différents coefficients par leur somme $s = \phi(1)$:

$$A(L) = \frac{\phi(L)}{\phi(1)} = a_0 + a_1 L + \dots + a_n L^n$$

$$\sum_{i=1}^n a_i = 1$$

Les coefficients a_i peuvent être assimilés à une distribution de probabilité définie sur les périodes $0, 1 \dots n$ (n fini ou infini). La fonction génératrice de la distribution de probabilité ainsi définie n'est autre que le polynôme $A(L)$. On peut alors déterminer les moments de la distribution ($a_0 \dots a_n$) en dérivant la fonction génératrice $A(L)$. Le délai moyen, qui représente le temps moyen de réaction de la variable y aux variations de la variable x , est égal au moment du premier ordre :

$$\theta = A'(1) = \frac{\Phi'(1)}{\Phi(1)}$$

La composition des retards précédemment évoquée peut être retrouvée en utilisant la propriété des fonctions génératrices des lois de probabilité (la fonction globale est égale au produit des fonctions composantes). Enfin, le délai moyen de la fonction globale de retards est égal à la somme des délais moyens. On a en effet :

$$\theta = \frac{[\Phi_1(L) \Phi_2(L)]'_{L=1}}{\Phi_1(1) \Phi_2(1)} = \frac{\Phi_1'(1)}{\Phi_1(1)} + \frac{\Phi_2'(1)}{\Phi_2(1)}$$

d'où :

$$\theta = \theta_1 + \theta_2$$

2.3. Distributions finies, infinies et formes autoregressives

Le modèle $y_t = \Phi(L) x_t$ qui exprime y en fonction des retards échelonnés de la variable x peut être transformé en une fonction à retards échelonnés de la variable y en utilisant la fonction inverse $\Phi^{-1}(L)$:

$$x_t = \Phi^{-1}(L) y_t$$

Lorsque $\Phi(L)$ est une distribution finie, son inverse est une distribution infinie et vice-versa. On peut ainsi transformer certaines distributions infinies en une distribution finie de la variable expliquée y_t , c'est-à-dire en une forme autoregressive finie. Ainsi en est-il de la distribution géométrique :

$$\Phi(L) = a(1 + \lambda L + \lambda^2 L^2 + \dots + \lambda^n L^n + \dots) = \frac{a}{1-\lambda L}$$

qui peut être transformée en une forme autoregressive du premier ordre (transformation de KOYCK) :

$$y_t - \lambda y_{t-1} = a x_t$$

Plus généralement, on peut supposer que la distribution de retards fait intervenir p valeurs retardées de la variable endogène (forme autoregressive d'ordre p) :

$$y_t = (b_1 y_{t-1} + b_2 y_{t-2} + \dots + b_p y_{t-p}) + (a_0 x_t + a_1 x_{t-1} + \dots + a_n x_{t-n})$$

ou encore :

$$y_t = W(L) y_t + P(L) x_t$$

La forme à retards échelonnés du modèle s'obtient en résolvant la relation par rapport à y et en développant en série la fonction rationnelle ainsi obtenue :

$$y_t = \frac{P(L)}{1 - W(L)} x_t$$

Si les racines de $[1 - w(z^{-1})]$ sont réelles, positives et inférieures à un, la distribution de retards échelonnés est positive et convergente (ses coefficients ainsi que leur somme tendent vers zéro lorsque les retards tendent vers l'infini). Lorsque les racines sont complexes ou négative, elle est oscillatoire.

JORGENSON [1966] a montré que tout modèle à retards échelonnés pouvait être approximé par une relation :

$$y_t = F(L) x_t$$

où $F(L)$ est une fraction rationnelle de l'opérateur L. Le degré du polynôme du dénominateur détermine l'ordre du modèle à retards échelonnés. Avant de passer à l'étude des distributions de retards les plus

couramment utilisées, nous allons comparer rapidement l'analyse des distributions de retards en temps discret que nous venons de développer avec l'analyse en temps continu que l'on utilise plus volontiers dans les modèles théoriques.

2.4. Temps discret, temps continu

L'analyse des retards en temps continu est antérieure à l'analyse en temps discret* et elle se prête mieux en général à l'étude de modèles théoriques.

On suppose que les grandeurs sont des fonctions continues et différentiables du temps. La distribution de retards discrète :

$$y_t = \sum_{i=0}^n w_i x_{t-i}$$

peut ainsi s'écrire en temps continu :

$$y_t = \int_0^{\tau} w(\theta) x(t-\theta) d\theta$$

Le coefficient de réaction à long terme (s) et le délai moyen de réaction sont respectivement :

$$S = \int_0^{\tau} w(\theta) d\theta \quad T = \frac{1}{S} \int_0^{\tau} w(\theta) \cdot \theta d\theta$$

L'équivalent de la forme autoregressive est, en analyse continue, une équation différentielle. Prenons par exemple la distribution géométrique ou distribution du premier ordre qui s'écrit, en notation discrète :

$$y_t - \lambda y_{t-1} = a x_t$$

./.

* cf. par exemple RGD ALLEN [] et les différents modèles de cycle.

En analyse continue, la différence $(y_t - y_{t-1})$ correspond à la dérivée* $\frac{dy}{dt}$, et l'opérateur L précédemment défini à l'opérateur $(1 - \frac{d}{dt})$. Avec ces notations, on retrouve les formules établies précédemment. La forme autoregressive du modèle est une équation différentielle qui s'écrit :

$$(1 - \lambda L) y_t = a x_t$$

ou encore :

$$(1 - \lambda) y_t - \lambda \frac{dy}{dt} = a x_t$$

Cette dernière relation peut être réécrite en faisant apparaître le délai moyen $T = \frac{\lambda}{1 - \lambda}$ et le coefficient à long terme $S = \frac{a}{1 - \lambda}$:

$$y_t - T \frac{dy}{dt} = S x_t$$

Son intégration donne la forme à retards échelonnés du modèle (retard simple exponentiel).

$$y_t = S \int_0^{\infty} \frac{1}{T} e^{-\frac{\theta}{T}} x(t - \theta) d\theta$$

En utilisant l'opérateur $D = \frac{d}{dt}$ ou $L = 1 - \frac{d}{dt}$, on peut retrouver toutes les propriétés précédemment évoquées. Nous ne nous étendrons pas sur l'analyse continue qui n'est guère utilisable dans les modèles empiriques, nous allons poursuivre l'étude des modèles à retards échelonnés en examinant les approximations les plus courantes c'est-à-dire d'une part l'utilisation de modèles autoregressifs du premier ou du second ordre, d'autre part l'approximation polynômiale (méthode d'ALMON).

3 - LES DISTRIBUTIONS INFINIES ET LEUR ESTIMATION SOUS FORME AUTO-REGRESSIVE (I) : PRECISION DES ESTIMATIONS PAR LES MOINDRES CARRÉS

Les distributions infinies utilisées en pratique correspondent toujours à une forme autoregressive simple, c'est-à-dire comportant peu de valeurs endogènes retardées. Les plus fréquentes sont les

./.

* Le lecteur désireux d'approfondir les retards échelonnés en analyse continue pourra se reporter à ALLEN [1969] p.104. Il faudra toutefois transcrire nos notations. Notre coefficient λ correspondant à $(1 - \lambda)$ dans l'ouvrage d'ALLEN, et ce dernier fait correspondre la dérivée

$D = \frac{d}{dt}$ à l'opérateur $y_{t+1} - y_t$ et non comme nous à $(y_t - y_{t-1})$.

distributions géométriques ou plus généralement les distributions du premier ordre et les distributions du second ordre.

La distribution géométrique correspond au modèle simple autoregressif :

$$(1) y = \lambda y_{-1} + a x,$$

dont la fonction de retards est la fraction rationnelle :

$$(1') \Phi(L) = \frac{a}{1-\lambda L}$$

Les distributions du premier ordre représentent à une première généralisation de la distribution géométrique, introduisant plusieurs retards échelonnés de la variable explicative :

$$(2) y = a_0 x + a_1 x_{-1} + \dots + a_n x_{-n} + \lambda y_{-1}$$

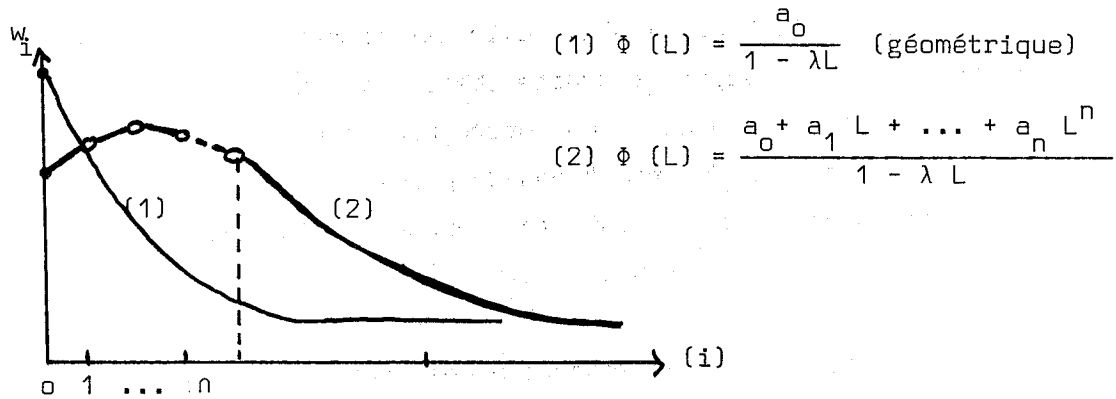
$$(2') \Phi(L) = \frac{a_0 + a_1 L + \dots + a_n L^n}{(1-\lambda L)}$$

Cette distribution, plus souple que la distribution géométrique, peut être considérée comme le produit d'une distribution discrète $(a_0 + a_1 L + \dots + a_n L^n)$ par une distribution géométrique $(\frac{1}{1-\lambda L})$. Le délai moyen de la distribution se calcule aisément en utilisant la composition des retards :

$$\theta = \frac{a_1 + 2 a_2 + \dots + n a_n}{a_0 + a_1 + \dots + a_n} + \frac{\lambda}{1-\lambda}$$

le premier terme correspondant à la distribution discrète, le seconde à la distribution géométrique. Le graphique (2) représente la forme générale des distributions du premier ordre. Elle correspond en pratique à une distribution discrète de 0 à n, et ensuite à une distribution géométrique.

GRAPHIQUE (2) - FORME GENERALE DES DISTRIBUTIONS DU 1er ORDRE



Sur données trimestrielles, et plus rarement sur données annuelles, on utilise parfois des distributions autoregressives du second ordre :

$$(3) y = a x + b y_{-1} + c y_{-2}$$

$$(3') \phi(L) = \frac{a}{1 - bL - cL^2}$$

ou plus généralement :

$$(4') \phi(L) = \frac{a_0 + a_1 L + \dots + a_n L^n}{1 - bL - cL^2}$$

Les distributions de retards ainsi obtenues sont convergentes et à coefficients positifs si les coefficients b et c vérifient les relations :

$$b^2 \geq -4c$$

$$0 < b < 2$$

$$|c| < 1$$

$$1 - b - c > 0$$

On peut représenter graphiquement la zone des valeurs acceptables de b et c (cf. GRILICHES [1967]). Elle est comprise entre la parabole $b^2 = -4c$, l'axe des ordonnées, et la droite $(b+c) = 1$ (zone hachurée sur le graphique-3). Une distribution du second ordre peut toujours être considérée comme le produit de deux distributions du premier ordre. Les paramètres λ_1 et λ_2 de ces distributions sont les racines de l'équation caractéristique :

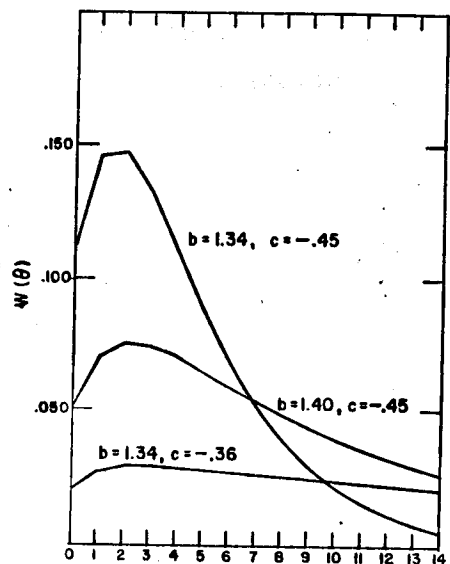
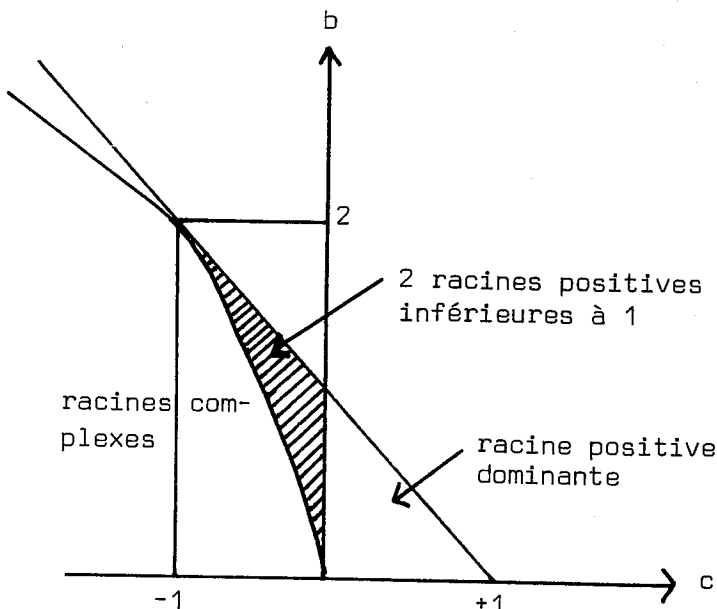
$$\lambda^2 - b\lambda - c = 0$$

La distribution des retards échelonnés s'exprime alors facilement en fonctions des paramètres λ_1 et λ_2 :

$$(3') \phi(L) = \frac{a}{(1-bL-cL^2)} = \frac{a}{(1-\lambda_1 L)(1-\lambda_2 L)} = \sum_{n=0}^{\infty} a \frac{(\lambda_1^{n+1} - \lambda_2^{n+1})}{(\lambda_1 - \lambda_2)} L^n$$

En pratique, elle ne diffère de la distribution géométrique que si b est supérieur à un (partie haute de la zone hachurée sur le graphique (3)).

GRAPHIQUE (3) - DISTRIBUTIONS DU SECOND ORDRE : ZONE DES VALEURS ACCEPTABLES DE b ET c ET FORME DES RETARDS ÉCHELONNES



Lorsque la théorie ne fixe pas une spécification précise pour les retards échelonnés, on a souvent le choix entre plusieurs spécifications alternatives. Nous verrons par exemple que l'on peut indifféremment représenter les délais de l'accélérateur flexible par une distribution du premier ordre ($\frac{a_0 + a_1 L}{1 - \lambda L}$) ou du second ordre ($\frac{a}{1 - bL - cL^2}$). De façon à caractériser l'incertitude qui affecte l'estimation des paramètres des modèles à retards échelonnés, nous allons étudier les estimations de ces différentes distributions en prenant divers exemples empiriques.

3.1. La distribution géométrique estimée sous forme autoregressive : quelques exemples

La distribution géométrique est, avec la méthode d'ALMON le modèle le plus répandu dans les études empiriques. La raison tient d'une part aux justifications théoriques de cette formulation (grandeur permanente, coûts d'ajustements), d'autre part à la simplicité de sa forme autorégressive qui se prête à une estimation facile par les moindres carrés ordinaires, sous réserve de l'hypothèse d'indépendance des erreurs. Nous évoquerons plus loin les problèmes que pose cette hypothèse, et les méthodes alternatives d'estimations de ces distributions. Nous allons pour l'instant étudier sur quelques exemples, la précision des estimations par les moindres carrés ordinaires.

Nous considérerons tout d'abord la fonction d'investissement liant le taux d'accumulation ($\frac{I}{K}$) au taux de profit retenu (π) :

$$\frac{I}{K} = \lambda \left(\frac{I}{K} \right)_{-1} + a \pi + d$$

Le tableau (1) présente l'estimation par les moindres carrés des coefficients λ et a sur série chronologique (une vingtaine d'observations) et sur données à deux dimensions (environ 200 observations).

TABLEAU (1) - ESTIMATIONS DE λ ET a

	λ	a	Source
Série chronologique 56-74 N = 19 observations	0,52 (3,5)	0,40 (2,7)	P.A.MUET [1978] page 21
18 branches x 11 années N = 198 observations	0,649 (15,0)	0,288 (6,9)	P.A.MUET [1976] page 21

La précision des estimations que l'on peut apprécier par la valeur des T de Student des coefficients est assez bien représentative de ce que l'on peut attendre d'estimations sur séries chronologiques ou sur données à deux dimensions pour d'autres fonctions (consommations, demande de monnaie ou demande d'emploi, encore que dans ce dernier cas la précision soit souvent plus faible).

La valeur du délai moyen ($\frac{\lambda}{1-\lambda}$) et celle du coefficient à long terme ($\frac{a}{1-\lambda}$) dépendent de façon cruciale de la précision (λ). On peut, en première approximation, calculer l'erreur totale sur le délai moyen par la formule d'approximation* :

$$ET = \frac{\sigma_{\lambda}}{(1-\lambda)^2}$$

Cette relation montre l'erreur totale est très importante lorsque le coefficient λ est proche de un. L'écart-type de λ est en effet souvent de

./.

* L'erreur totale est la racine carrée de l'écart quadratique moyen de l'estimation qui est égal à la somme de la variance et du carré du biais. Dans le cas d'un rapport $\frac{1}{V}$, on peut, en négligeant les moments d'ordre supérieur à 2, approximer l'espérance mathématique et l'erreur totale par :

$$E\left(\frac{1}{V}\right) = \frac{1}{E(V)} + \frac{\sigma_V^2}{[E(V)]^3} \quad \text{et :} \quad ET = \frac{\sigma_V}{[E(V)]^2}$$

l'ordre de 0,1 en moyenne, sur série chronologique. Avec cette hypothèse, la précision dans l'estimation du délai moyen est très médiocre. Comme le montre le tableau (2). Le coefficient λ est significativement différent de zéro et de un lorsqu'il est compris entre 0,3 et 0,7 et dans toute cette zone, l'erreur totale sur le délai moyen est presque égale à la moitié de la valeur du délai moyen.

TABLEAU (2) - DELAI MOYEN ET ERREUR TOTALE SUR LE DELAI MOYEN
POUR DIVERSES VALEURS DE λ LORSQUE $\sigma \lambda = 0,1$

λ	N.S.D de 0		S.D. de 0 et 1					N.S.D de 1	
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
θ années	0,10	0,25	0,4	0,7	1,0	1,5	2,3	4,0	9,0
ET (θ) années	0,10	0,15	0,2	0,3	0,4	0,6	1,1	2,5	10,0

Dans le modèle considéré, l'écart-type de λ est égal à 0,14 sur série chronologique et 0,04 sur données sectorielles-chronologiques. Les délais moyens et les erreurs sont respectivement 1,1 ans et 0,6 ans sur série chronologique, 1,8 ans et 0,3 ans sur données sectorielles chronologiques.

Ce que nous avons dit du délai moyen s'applique également au coefficient à long terme. Pour ce dernier, il faut néanmoins tenir compte de la corrélation entre les estimations des coefficients $\hat{\alpha}$ et $\hat{\lambda}$. Si celle-ci est négative, comme c'est souvent le cas, la précision dans l'estimation du coefficient moyen sera meilleure que ce que suggère le simple examen des écarts-types des coefficients, car il y aura en quelque sorte une compensation entre $\hat{\alpha}$ et $(\frac{1}{1-\lambda})$. Pour étudier de façon plus rigoureuse la précision dans l'estimation du coefficient à long terme, on peut utiliser l'intervalle de confiance du couple* ($\hat{\alpha}$, $\hat{\lambda}$) qui est une ellipse définie par :

* On peut utiliser aussi la méthode suggérée par FIELLER et appliquée au délai moyen (voir paragraphe 34).

$$[(a-\hat{a}), (\lambda-\hat{\lambda})] \begin{bmatrix} \sigma_a^2 & \sigma_{a\lambda} \\ \sigma_{a\lambda} & \sigma_\lambda^2 \end{bmatrix} \begin{bmatrix} a - \hat{a} \\ \lambda - \hat{\lambda} \end{bmatrix} = 2 F_{2, n-k}(\alpha)$$

où $F_{2, n-k}(\alpha)$ est la valeur critique de $F_{2, n-k}$ pour le seuil de confiance α (5 % dans le cas présent).

Le graphique (4) représente cette ellipse pour l'estimation la plus précise (données sectorielles-chronologiques), ainsi que l'intervalle de confiance du couple $(\hat{\lambda}, \frac{\hat{a}}{1-\hat{\lambda}})$ (partie droite du graphique 4). Cet intervalle de confiance montre que le coefficient à long terme du profit, dont la valeur estimée est 0,82, n'est pas significativement différent au seuil de probabilité de 5 % de toute valeur comprise entre 0,57 et 1,02. On remarquera que cet intervalle de confiance est décalé vers le bas, par rapport à la valeur estimée du coefficient à long terme.

Ce résultat montre que même dans le cas d'une estimation précise des coefficients de la forme autoregressive du modèle, le coefficient à long terme et le délai moyen sont estimés avec une assez grande incertitude. Aussi ne doit-on pas hésiter, lorsque la théorie le suggère, à fixer à priori la valeur du coefficient à long terme. Ce sera par exemple le cas de l'accélérateur flexible :

$$(1) \quad \left(\frac{I}{K}\right) = \lambda \left(\frac{I}{K}\right)_{-1} + a \frac{Q}{K} + d,$$

ou celui des délais d'ajustement de l'emploi :

$$(1') \quad \text{Log } N = \lambda' \text{Log } N_{-1} + a' \text{Log } N^*$$

avec $N^* = \bar{w} e^{-\gamma t} \cdot Q$

GRAPHIQUE (4) - INTERVALLE DE CONFIANCE DU COEFFICIENT A LONG TERME
(MODELE DE PROFIT SUR DONNEES SECTORIELLES CHRONOLOGIQUES)

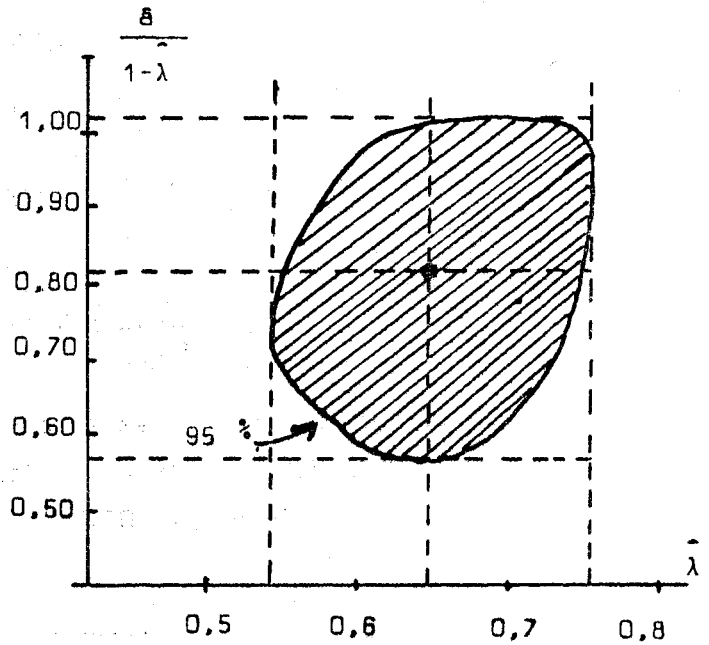
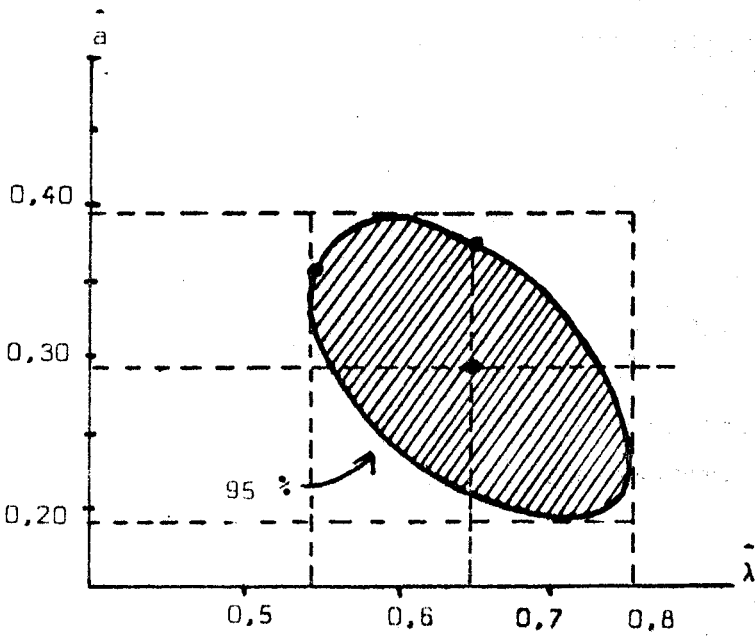
a - Intervalle de confiance de $(\hat{a}, \hat{\lambda})$

$$\hat{\lambda} = 0,649 \quad \sigma_{\lambda} = 0,043$$

$$\hat{a} = 0,288 \quad \sigma_{\hat{a}} = 0,041$$

$$\sigma_{a\lambda} = -0,95 \cdot 10^{-3}$$

b - Intervalle de confiance de $(\frac{\hat{a}}{1-\hat{\lambda}}, \hat{\lambda})$



Dans les deux modèles, le coefficient à long terme ($\frac{a}{1-\lambda}$) doit être théoriquement égal à un. L'estimation de ces modèles sur série chronologique annuelle par les moindres carrés ordinaires est très imprécise et, dans les deux cas, le coefficient à long terme n'est pas significativement différent de un. L'estimation des mêmes modèles sous la contrainte d'un coefficient à long terme égal à un (c'est-à-dire $a=1-\lambda$), peut être obtenue en appliquant les moindres carrés aux équations transformées

$$(a) \quad \left[\frac{I}{K} - Q \right] = \lambda \left[\left[\frac{I}{K} - Q \right]_{-1} + d \right]$$

$$(2') \quad \text{Log} \left(\frac{N}{N_{-1}} \right) = (1-\lambda) \text{Log} \left(\frac{N^*}{N_{-1}} \right)$$

On obtient par exemple pour l'estimation de l'accélérateur flexible dans la branche "bien de consommation" (Source P.A. MUET, P. ZAGAME [1976] p. 107).

	Modèle non contraint		Modèle contraint ($a=1-\lambda$)
	λ	a	λ
Coefficient estimé (écart-type)	0,885 (0,071)	0,205 (0,046)	0,822 (0,037)

L'estimation sous contrainte diminue l'écart-type du terme autoregressif (λ) de moitié, et l'erreur totale sur le délai moyen passe de 7 ans à 1 an environ. Le gain en précision est encore plus important pour la fonction d'ajustement de l'emploi où, en raison de la colinéarité entre N_{-1} et N^* , les coefficients λ' et a' ont des écarts-types qui représentent plus de 50 % de la valeur du coefficient. Lorsqu'on contraint le coefficient à long terme a être égal à un, l'estimation du coefficient (λ') devient précise comme le montre les estimations du cycle de productivité des trois branches industrielles du modèle DMS (cf. D. FOUQUET ...[1978] page 63).

	Intermédiaires	Equipement	Consommation
(1-λ')	0,28	0,33	0,43
Ecart-type	(0,05)	(0,03)	(0,05)

3.2. Distributions du premier ordre : l'accélérateur flexible sur données annuelles

De nombreuses estimations de l'accélérateur flexible sur données annuelles on conduit à retenir, pour représenter les délais de l'effet d'accélération, une distribution de premier ordre :

$$\phi(L) = \frac{a_0 + a_1 L}{1 - \lambda L}$$

Ainsi spécifié, le modèle peut être estimé sous forme autoregressive :

$$\left(\frac{I}{K}\right) = \lambda \left(\frac{I}{K}\right)_{-1} + a_0 Q + a_1 Q_{-1} + d$$

si l'on suppose que les erreurs de la forme autoregressive sont sans corrélation. Le tableau (3) ci-après présente trois estimations réalisées sur les données annuelles de la comptabilité nationale. Les deux premières correspondent à un ensemble de données sectorielles-chronologiques comportant respectivement 120 et 234 observations. La troisième a été obtenue sur série chronologique et nous en donnons deux estimations qui mettent en évidence une assez grande instabilité des coefficients (le simple examen de l'évolution de $\frac{I}{K}$ montre que la période 53-56 est assez particulière).

TABEAU (3) - TROIS ESTIMATIONS DE L'ACCELERATEUR FLEXIBLE
SUR DONNEES ANNUELLES*

	λ	a_0	a_1	Nature de l'estimation N = nombre d'observations
Investissement Industriel Source : V. THOLLON- POMMEROL, E. MALINVAUD [1971]	0,73 (0,05)	0,113 (0,026)	0,093 (0,027)	MCO sur données sectorielles-chronologiques (56-70):N=8x15=120
Investissement Industriel Source : P.A. MUET [1978]	0,682 (0,042)	0,095 (0,026)	0,096 (0,026)	MCO sur données sectorielles-chronologiques (61-73)N=18x13=234
Investissement en équipements des entreprises non financières	0,610 (0,068)	0,260 (0,039)	0,127 (0,047)	N = 22 (53-74)
	0,488 (0,099)	0,237 (0,041)	0,105 (0,048)	N = 19 (56-74)

* (N) représente le nombre d'observations, les écarts-types figurent entre parenthèses.

On remarquera que les estimations sur série chronologique, qui comportent un nombre réduit d'observations sont, comme précédemment, moins précises que les estimations sur données croisées : les écarts-types sont en effet environ une fois et demi à deux fois plus élevés sur série chronologique que sur données sectorielles-chronologiques. Nous allons étudier de façon plus détaillée l'estimation du modèle sur série chronologique qui nous permettra en outre une comparaison avec les distributions du second ordre.

Les coefficients à long terme et les délais moyens estimés diffèrent sensiblement dans les deux périodes, mais l'estimation des écarts-types des coefficients que l'on peut obtenir de façon approchée à partir des matrices des variances-covariances de (a_0, a_1, λ) montre que la différence n'est pas significative au seuil habituel (5 %).

TABLEAU (4) - VALEUR DES COEFFICIENTS A LONG TERME ET
DES DELAIS MOYENS DES ESTIMATIONS SUR SERIE
CHRONOLOGIQUE

	Coefficient à long terme		Délais moyens (années)	
	CLT = $(a_0 + a_1) / (1-\lambda)$		$\theta = a_1 / (a_0 + a_1) + \lambda / (1-\lambda)$	
	53-74	56-74	53-74	56-74
Coefficient (écart-type)	0,99 (0,15)	0,67 (0,15)	1,89 (0,41)	1,25 (0,34)

Le tableau (4) montre également que le coefficient à long terme n'est pas significativement différent de un. Ce résultat est évident sur la période 1953-1974 tandis que sur la période 1956-1974, la valeur un est la limite supérieure de l'intervalle de confiance (en outre, l'écart-type sous-estime l'erreur totale puisqu'il ne tient pas compte du biais). On peut donc estimer la spécification théorique de l'accélérateur flexible qui impose la contrainte :

$$a_0 + a_1 = 1 - \lambda$$

Mais alors que nous avons pu dans le cas d'une distribution géométrique appliquer les moindres carrés ordinaires au modèle transformé, il faut, dans le cas présent présent estimer un modèle non linéaire en ses coefficients.

Posons en effet $a_0 = \alpha (1-\lambda)$ $a_1 = (1-\alpha) (1-\lambda)$, on obtient

$$\frac{I}{K} = (1-\lambda) \alpha \overset{0}{Q} + (1-\lambda) (1-\alpha) \overset{0}{Q}_{-1} + \lambda \left(\frac{I}{K}\right)_{-1} + d + \varepsilon$$

Si les erreurs (ε) suivent une loi normale entrée réduite, l'estimation du maximum de vraisemblance équivaut à la minimisation de la variance résiduelle. On obtient par exemple dans le cas présent :

TABLEAU (5) - ESTIMATION SOUS CONTRAINTE DE
L'ACCELERATEUR FLEXIBLE (MINIMISATION DE
LA VARIANCE RESIDUELLE PAR ITERATIONS)

	α	λ	θ -Délai moyen
Période 1953-1974	0,672 (0,086)	0,612 (0,053)	1,90 (0,31)
Période 1956-1974	0,709 (0,100)	0,618 (0,063)	1,90 (0,38)

Les deux estimations du modèle deviennent cette fois-ci très proche et la précision dans l'estimation des retards échelonnés s'améliore. En utilisant comme précédemment l'intervalle de confiance du couple (α, λ) , on peut, comme nous le verrons au paragraphe 3.4, représenter la forme de différentes distributions de retards situées sur la frontière de l'intervalle de confiance (partie haute du graphique 5).

3.3. Distributions du second ordre

Les distributions du second ordre ont souvent été utilisées pour estimer les distributions de retards sur série trimestrielle. Il est rare en revanche que celles-ci donnent des coefficients significatifs sur données annuelles. Le tableau (6) ci-après indique trois estimations de distributions du second ordre. La première correspond à une fonction d'investissement estimée sur les séries trimestrielles de l'industrie manufacturière américaine par GRILICHES et WALLACE [1965] :

$$I_t = 0,0126 K_t + 0,0111 q_{t-2} - 0,581 r_{t-2} + 1,150 I_{t-1} - 0,331 I_{t-2}$$

(0,0052) (0,0035) (0,222) (0,117) (0,109)

K_t est le stock de capital en début de trimestre (t)
 I_t l'investissement net au cours du trimestre (t)
 q_{t-2} un indice des "valeurs industrielles" au trimestre (t-2)
 r_{t-2} le taux d'intérêt au trimestre (t-2)

Les deux autres, à l'accélérateur flexible :

$$\frac{I}{K} = a Q + b \left(\frac{I}{K}\right)_{-1} + c \left(\frac{I}{K}\right)_{-2} + d,$$

estimé d'une part sur série chronologique 1953-1976, d'autre part sur données sectorielles-chronologiques (mêmes données qu'au paragraphe précédent).

Les écarts types des coefficients b et c sont sensiblement du même ordre de grandeur (0,1) dans les estimations sur série chronologique annuelle ou trimestrielle et plus faibles sur données sectorielles x chronologiques (0,06), mais dans ce dernier cas le coefficient (c) n'est pas significativement différent de zéro, c'est-à-dire que la distribution n'est pas significativement différente de la distribution géométrique. Ces trois estimations conduisent en outre à des distributions de retards convergentes et à coefficients positifs car le couple de coefficients (\hat{b} , \hat{c}) se trouve dans tous les cas à l'intérieur de la zone des valeurs acceptables représentée sur les graphiques (3) et (5). Nous allons étudier plus en détail l'estimation de l'accélérateur flexible sur série chronologique annuelle 1953-1974, en comparant notamment les distributions du second ordre et du premier ordre.

TABLEAU (6) - TROIS ESTIMATIONS D'UNE DISTRIBUTION DU SECOND
ORDRE APPLIQUEES AUX DELAIS DE LA FONCTION
D'INVESTISSEMENT

	\hat{b}	\hat{c}	\hat{a}	Nature de l'estimation
Investissement Industriel (source GRILICHES, WALLACE [1965])	1,150 (0,117)	- 0,331 (0,110)		MCO sur série trimestrielle 48-62: N = 62
Investissement en équipement des ENF	0,896 (0,100)	- 0,202 (0,100)	0,246 (0,042)	MCO sur série chronologique annuelle: 53-76 N = 22
Investissement industriel (source P.A. MUET [1978])	0,723 (0,063)	- 0,040 (0,064)	0,094 (0,028)	MCO sur données sectorielles chronologiques : N = 18 x 13 = 234

3.4. Comparaison des distributions du premier et du second ordre : accélérateur flexible estimé sur série chronologique annuelle (1953-1974)

Si l'on ne tient pas compte de la contrainte d'un coefficient à long terme égal à un, la distribution des retards échelonnés dépend dans les deux cas de trois paramètres :

- a_0, a_1, λ pour la distribution du 1er ordre
- a, b, c pour la distribution du second ordre

Ces deux distributions peuvent être caractérisées par leur délai moyen :

$$\theta_1 = \frac{a_1}{a_0 + a_1} + \frac{\lambda}{1-\lambda} \quad \text{1er ordre ;}$$

$$\theta_2 = \frac{1}{1-b-c} \quad \text{2ème ordre ;}$$

et par leur coefficient à long terme :

$$CLT_1 = \frac{a_0 + a_1}{1-\lambda} \quad \text{1er ordre,}$$

$$CLT_2 = \frac{a}{1-b-c} \quad \text{2ème ordre.}$$

Le Tableau (7) indique la valeur de ces coefficients, ainsi que celle de l'écart-type et du T de Student.

Les coefficients à long terme et les délais moyens ne sont pas significativement différents mais la précision des estimations est meilleure avec une distribution du premier ordre. Enfin la variance résiduelle (somme des carrés des résidus $\sum e^2$) montre que l'ajustement de la distribution du premier ordre est légèrement meilleur que celui de la distribution du second ordre.

TABLEAU (7) : ACCELERATEUR FLEXIBLE - COMPARAISON DES DISTRIBUTIONS DU PREMIER ET DU SECOND ORDRE

	Coefficient à LT		Délai moyen (années)		$\sum e^2 \cdot 10^{-3}$	
	CLT ₁	CLT ₂	θ_1	θ_2	(1)	(2)
Coefficient	0,99	0,80	1,89	1,60	0,105	0,121
Ecart-type (σ)	(0,15)	(0,20)	(0,41)	(0,67)		
(T)Student	(6,5)	(4,0)	(4,6)	(2,4)		

Lorsqu'on contraint le coefficient à long terme à être égal à un, les distributions de retards ne dépendent plus que de deux coefficients :

$$\begin{array}{lll}
 - \hat{\alpha} = 0,672 & \hat{\lambda} = 0,612 & \phi(L) = \frac{(1-\lambda) [\alpha + (1-\alpha)L]}{(1-\lambda L)} \quad \text{1er ordre} \\
 (0,086) & (0,053) & \\
 - \hat{b} = 0,901 & \hat{c} = -0,167 & \phi(L) = \frac{1-b-c}{1-bL-cL^2} \quad \text{2ème ordre} \\
 (0,099) & (0,090) &
 \end{array}$$

La précision dans l'estimation des coefficients s'améliore, mais assez peu cependant pour la distribution du second ordre et surtout, la diminution du coefficient c en valeur absolue rend sa valeur estimée non significativement différente de zéro. Pour juger de la précision de ces deux estimations, et pour comparer la forme des retards échelonnés on a tracé les intervalles de confiance des couples $(\hat{\alpha}, \hat{\lambda})$ et (\hat{b}, \hat{c}) et représenté d'une part la forme des retards échelonnés des distributions estimées (courbes en trait plein), d'autre part la forme de différentes distributions de retards correspondent à des points B, C, D ... situés aux limites de l'intervalle de confiance à 95 % (courbes en pointillé sur le graphique). La comparaison des deux distributions estimées (courbes A en trait plein) montre tout d'abord que l'échelonnement des retards est pratiquement le même, que la distribution soit du premier ou du deuxième ordre, ce que traduit également la valeur respective des délais moyens (1,9 ans pour la distribution du 1er ordre, 2,1 ans pour la distribution du second ordre). La précision est en revanche bien meilleure pour la distribution du 1er ordre. Si l'on trace en effet pour celle-ci les distributions de retards correspondant à deux points extrêmes de l'intervalle de confiance (B et C), on obtient des courbes assez proches de la distribution estimée, l'incertitude étant maximale pour un délai d'un an. En particulier, l'intervalle de confiance ne coupe pas la droite $\alpha=1$, ce qui montre que la distribution est significativement différente d'une distribution géométrique*.

La même analyse appliquée à la distribution du second ordre montre au contraire que l'échelonnement des retards est estimé avec une moins bonne précision. L'intervalle de confiance n'est pas entièrement inclus dans la zone des valeurs acceptables (hachurée sur le graphique) c'est-à-dire qu'il comprend des distributions de retards à coefficients

./.

* le test correspondant à un intervalle de confiance à 95 % est plus sévère que le test de STUDENT habituel au seuil de 5 % (Cf. Annexe).

positifs puis négatifs (cas du point D qui correspond à des racines λ_1 et λ_2 complexes, ou des points situés à droite du point B). On remarque en outre que la distribution estimée n'est pas significativement différente de la distribution géométrique (point B) et de la distribution de Pascal du second ordre (point C). Les trois distributions (B, C, D) situées sur la limite de l'intervalle de confiance mettent bien en évidence l'incertitude qui affecte l'estimation des retards échelonnés par une distribution du second ordre*. Celle-ci apparaît également lorsqu'on calcule l'incertitude qui affecte les délais moyens. Les programmes de régression standard (par exemple la version 3.4 de TSP) fournissent une estimation des écarts-types de différentes fonctions des coefficients estimés des régressions. Dans le cas présent on obtient par exemple pour les délais moyens (en années) :

$$\begin{array}{lll} \hat{\theta}_1 = 1,9 & \hat{\sigma}_{\theta 1} = 0,32 & \text{1er ordre} \\ \hat{\theta}_2 = 2,1 & \hat{\sigma}_{\theta 2} = 0,53 & \text{2ème ordre} \end{array}$$

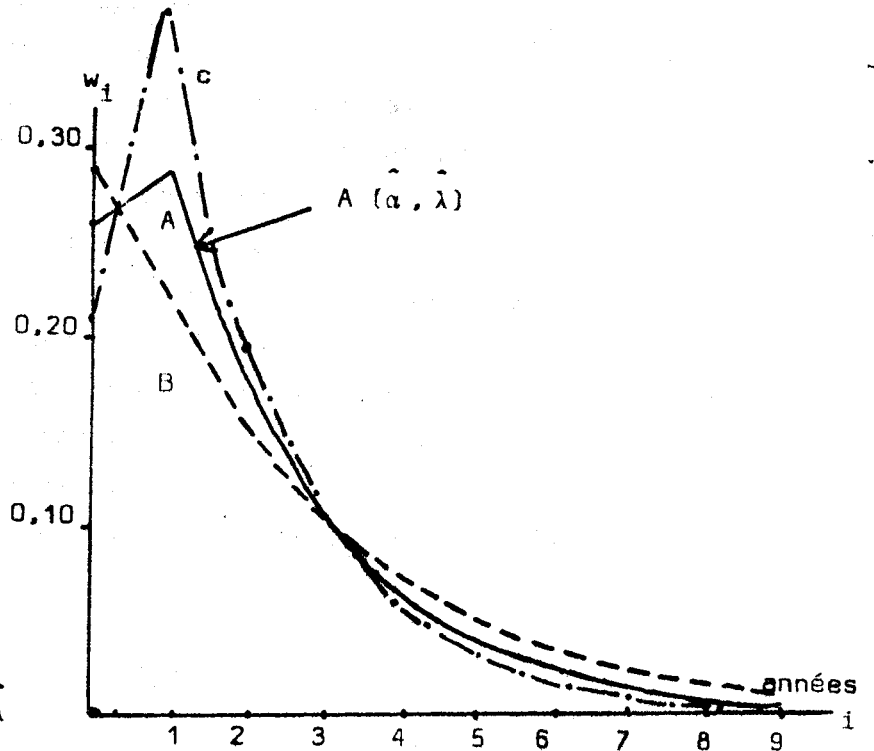
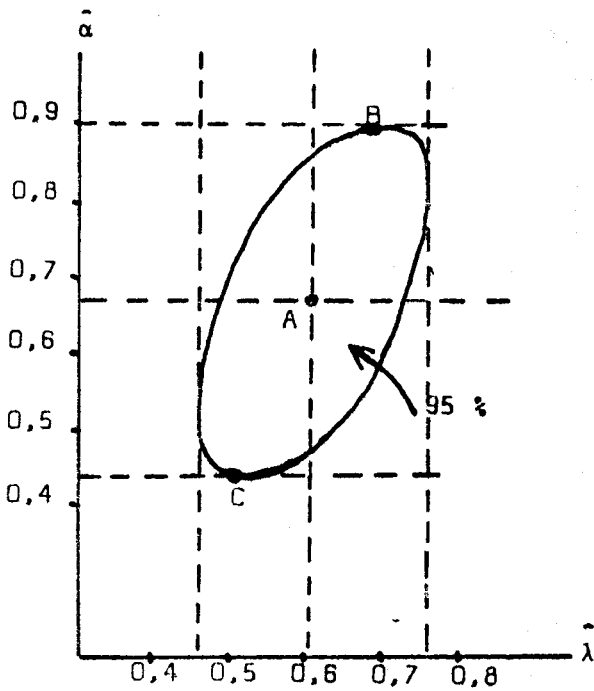
Les valeurs des écarts-types, obtenues par des approximations linéaires des relations définissant les délais moyens ne permettent pas de construire un intervalle de confiance par les méthodes habituelles (par exemple $\hat{\theta} \pm 2 \hat{\sigma}_{\theta}$ en prenant $T_{\alpha} = 2$), car l'estimation du délai moyen est biaisée, et l'intervalle de confiance n'est pas centré sur $\hat{\theta}$.

./.

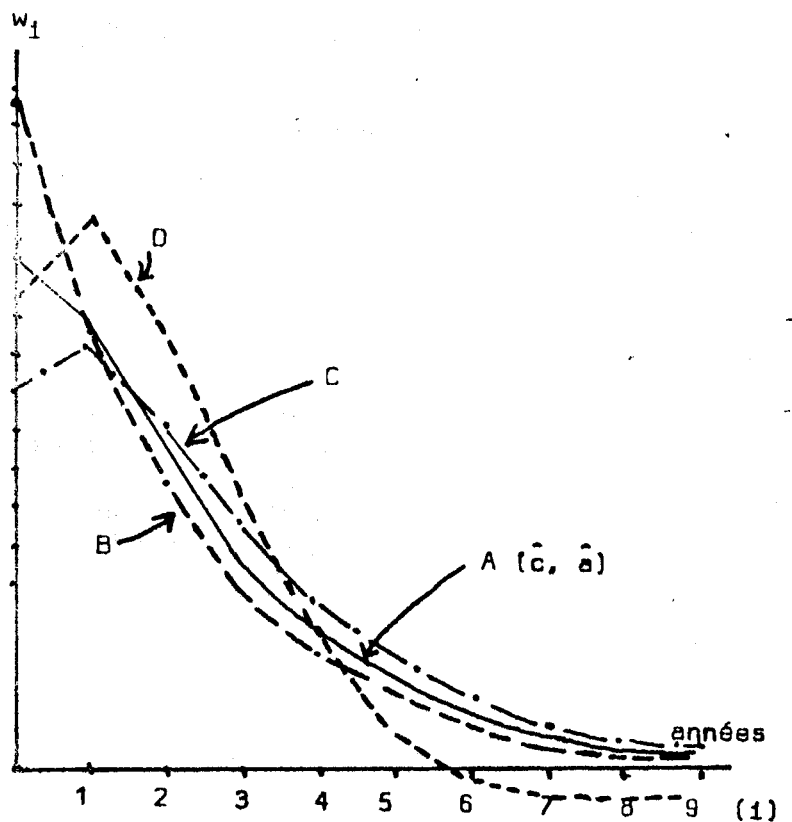
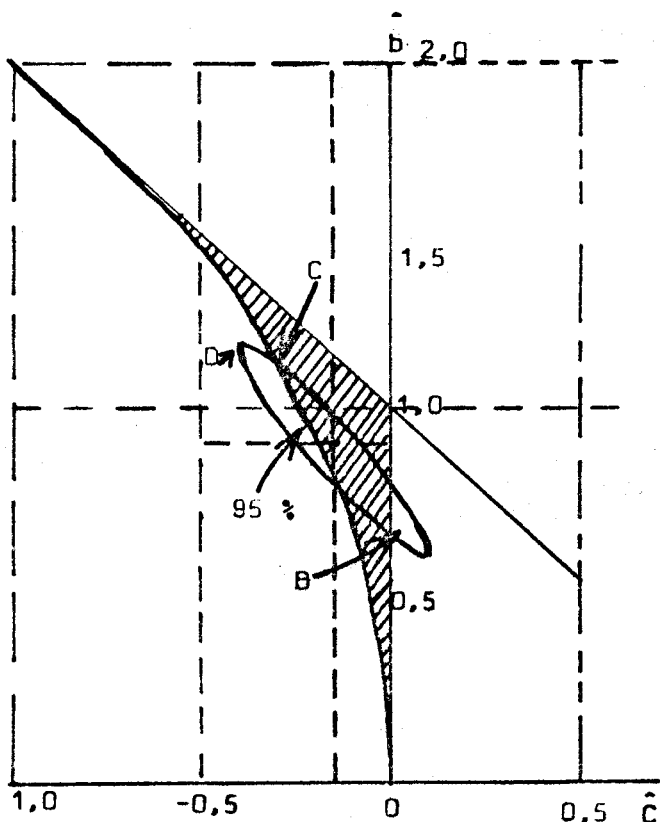
* Cette analyse est inspirée de Z. GRILICHES [1967]. Le lecteur pourra notamment consulter les graphiques identiques, correspondant à la fonction d'investissement GRILICHES-WALLACE, pages 30 à 32 de l'article de GRILICHES.

GRAPHIQUE (5) - DISTRIBUTIONS DU PREMIER ET DU DEUXIEME ORDRE ; INTERVALLE DE CONFIANCE A 95 % ET FORME DES RETARDS ECHELONNES CORRESPONDANTS A DIFFERENTS POINTS DE LA FRONTIERE

a - Distribution du 1er ordre



b - Distribution du second ordre



Pour la distribution géométrique simple où θ ne dépend que d'un paramètre (λ), nous avons indiqué des formules d'approximation du biais et de l'erreur totale. Celles-ci peuvent être utilisées également pour la distribution du premier ordre ($\hat{\theta} = \hat{\alpha} + \frac{\hat{\lambda}}{\hat{\alpha}}$) en tenant compte en outre de la covariance entre $\hat{\lambda}$ et $\hat{\alpha}$. Pour la distribution du second ordre où θ est le rapport de deux fonctions linéaires ($H = \hat{b} + 2\hat{c}$) et ($G = 1 - \hat{b} - \hat{c}$), on peut utiliser une méthode suggérée par FIELLER* et appliquée à la distribution du second ordre par GRILICHES [1967]. Si l'on définit en effet la variable $Z = H - G\theta$ d'espérance mathématique zéro et de variance :

$$\sigma_Z^2 = \sigma_H^2 + \theta^2 \sigma_G^2 - 2\theta \sigma_{HG}$$

on peut, sous l'hypothèse habituelle de normalité, construire un intervalle de confiance (θ', θ'') pour θ en résolvant l'équation du second degré :

$$(T_\alpha \cdot \sigma_Z)^2 = (H - G\theta)^2$$

ou encore :

$$\theta^2 (G^2 - T_\alpha^2 \cdot \sigma_G^2) - 2\theta (HG - T_\alpha^2 \cdot \sigma_{HG}) + H^2 - T_\alpha^2 \sigma_H^2 = 0$$

où T_α est la valeur du T de Student au seuil α (en pratique, au-delà de 20 observations $T_{0,05}$ est de l'ordre de 2).

Dans le cas présent on a :

$$\begin{aligned} H &= 0,567 & \sigma_H^2 &= 0,9149 \cdot 10^{-2} & \sigma_{GH} &= -0,112 \cdot 10^{-2} \\ G &= 0,266 & \sigma_G^2 &= 0,1276 \cdot 10^{-2} \end{aligned}$$

d'où l'on déduit $\theta' = 1,3$ années, $\theta'' = 3,3$ années.

L'intervalle de confiance ainsi obtenu diffère peu de l'intervalle que l'on aurait pu construire, de façon approximative, à partir de l'écart type estimé par le programme TSP :

- Intervalle de confiance réel : $\hat{\theta} = 2,1 \begin{cases} +1,2 \\ -0,8 \end{cases}$ années
- Intervalle "approximatif" : $\hat{\theta} = 2,1 \pm 1,1$ années

./.

* On en trouvera une présentation dans FULLER [1962] et dans GRILICHES [1967].

Comme pour la distribution du premier ordre, ce résultat est généralement vérifié lorsque le dénominateur du délai moyen ($G = 1 - \lambda$ pour le 1er ordre, $G = 1 - b - c$ pour le second) est suffisamment éloigné de zéro, c'est-à-dire lorsque la somme des termes autoregressifs est suffisamment et significativement inférieure à un*. La valeur centrale de l'intervalle de confiance est en effet :

$$\theta_n = \frac{\theta' + \theta''}{2} = \frac{GH - T_\alpha \cdot \sigma_{GH}}{G^2 - T_\alpha^2 \sigma_G^2}$$

Et cette valeur centrale diffère peu du délai moyen estimé :

$$\hat{\theta} = \frac{H}{G}$$

lorsque le T de Student de G est élevé.

Dans l'estimation de Z GRILICHES, T.D. WALLACE** où $G = 0,18$ et $\sigma_G^2 = 0,162 \cdot 10^{-2}$, l'intervalle de confiance du délai moyen est beaucoup plus décentré :

$$\hat{\theta} = 2,7 \left\{ \begin{array}{l} + 3,3 \\ - 1,6 \end{array} \right. \text{ trimestres}$$

3.5. Modèle autoregressif à plusieurs variables

Lorsque le modèle comporte plusieurs variables à retards échelonnés :

$$y_t = \phi_1(L) x_t + \phi_2(L) z_t$$

où $\phi_1(L)$ et $\phi_2(L)$ sont par exemple des distributions du premier ou du second ordre, l'estimation du modèle sous forme autoregressive devient beaucoup plus délicate. Même si l'on néglige le fait que les erreurs de la forme autoregressive ne sont pas indépendantes, on est conduit généralement à estimer des modèles linéaires comportant des contraintes sur les coefficients.

./.

* Ce résultat s'appliquerait à une distribution du même ordre ($G=1-b_1 \dots -b_n$).

** Comme les variables exogènes sont décalées de 2 trimestres, il faudrait ajouter 2 trimestres pour obtenir le délai moyen réel.

Considérons par exemple le cas de deux distributions du premier ordre :

$$y_t = \frac{a_0 + a_1 L}{1 - \lambda_1 L} x_t + \frac{b_0 + b_1 L}{1 - \lambda_2 L} z_t + \varepsilon_t,$$

qui conduit à la forme autoregressive :

$$y_t = (\lambda_1 + \lambda_2) y_{t-1} - \lambda_1 \lambda_2 y_{t-2} + a_0 x_t + (a_1 - \lambda_2 a_0) x_{t-1} - a_1 \lambda_2 x_{t-2} + b_0 z_t + (b_1 - \lambda_1 b_0) z_{t-1} - b_1 \lambda_1 z_{t-2} + u_t$$

Si l'on suppose que les erreurs (u_t) de la forme autoregressive sont indépendantes, il est possible d'estimer les paramètres de la forme autoregressive en minimisant la variance résiduelle. On peut soit procéder par "balayage" sur diverses valeurs de (λ_1, λ_2) en appliquant les moindres carrés à des variables transformées :

$$X_t = (1 - \lambda_2 L) x_t, Z_t = (1 - \lambda_1 L) z_t, Y_t = (1 - \lambda_2 L) (1 - \lambda_1 L) y_t,$$

soit pratiquer par itérations (c'est la méthode utilisée dans les programmes courants d'estimation). A titre d'exemple, nous présentons ci-après deux estimations du modèle accélérateur-profit sur données sectorielles-chronologiques (144 observations pour le modèle [1], 162 pour le modèle [2]). L'étude des délais d'ajustement de l'accélérateur flexible et du modèle de profit ont montré que l'on pouvait retenir une distribution du 1er ordre pour le premier et une distribution géométrique pour le second. Le modèle complet peut donc être spécifié comme suit :

$$[1] \quad \frac{I}{K} = \frac{a_0 + a_1 L}{1 - \lambda_1 L} \frac{Q}{Q} + \frac{b}{1 - \lambda L} \pi + d$$

o

Q est le taux de croissance de la valeur ajoutée

π est le taux de profit.

L'estimation des moindres carrés est obtenue après 6 itérations (méthode de GAUSS) sur la forme autoregressive du modèle (144 observations) La variance résiduelle ($\sum e^2$) a les valeurs successives :

itérations	1	2	3	4	5	6
$\Sigma e^2 \cdot 10^{-4}$	560,75	328,81	326,37	326,11	326,09	326,08

La valeur estimée des coefficients ainsi que les écarts-types sont indiqués dans le tableau ci-après :

Coefficient (écart- type)	\hat{a}_0	\hat{a}_1	$\hat{\lambda}_1$	\hat{b}	$\hat{\lambda}_2$	Coefficient à LT		Délai moyen	
						Q	π	Q	π
						0,045 (0,031)	0,077 (0,028)	0,62 (0,11)	0,439 (0,089)

Ce tableau montre que la distribution des retards des profits est beaucoup plus courte que celle de l'effet d'accélération, le coefficient $\hat{\lambda}_2$ n'étant pas en outre significativement différent de zéro. Une distribution discrète peut donc être retenue pour représenter les délais de réaction de l'investissement aux variations du taux de profits :

$$\frac{I}{K} = \frac{a_0 + a_1 L}{1 - \lambda_1 L} Q + (b_0 + b_1 L) \pi + d$$

La convergence est obtenue par la méthode de GAUSS après 4 itérations (162 observations) :

itérations	1	2	3	4
$\Sigma e^2 \cdot 10^{-4}$	1198,3	420,49	376,88	376,87

La valeur estimée des coefficients est indiquée ci-après :

	\hat{a}_0	\hat{a}_1	$\hat{\lambda}_1$	\hat{b}_0	\hat{b}_1	Coefficient à LT		Délai moyen	
						σ_Q	π	σ_Q	π
Coefficient (écart-type)	0,045 (0,030)	0,059 (0,030)	0,75 (0,05)	0,408 (0,093)	0,317 (0,131)	0,42 (0,17)	0,73 (0,17)	3,5 (0,9)	0,44 (0,11)

4 - LES DISTRIBUTIONS INFINIES ET LEUR ESTIMATION SOUS FORME AUTO-REGRESSIVE (II) : PROBLEMES D'IDENTIFICATION ET D'ESTIMATION

Dans la précédente section, nous avons estimé les différentes distributions de retards, en appliquant les moindres carrés à la forme autoregressive du modèle. Cette méthode qui a le mérite d'être simple, n'est justifiée que si les erreurs (u_t) de la forme autoregressive sont indépendantes :

$$y_t = P(L) x_t + W(L) y_t + u_t$$

Or, la validité de cette hypothèse doit être examinée avec soin car elle affecte de façon cruciale les propriétés de l'estimateur des moindres carrés, et de plus, on peut penser à contrario que ce sont les erreurs (ϵ_t) du modèle à retards échelonnés qui sont indépendantes :

$$y_t = \frac{P(L)}{1 - W(L)} x_t + \epsilon_t$$

auquel cas celles de la forme autoregressive sont liées :

$$u_t = \epsilon_t - W(L) \epsilon_{t-1}$$

Les biais dans l'estimation des modèles autoregressifs par les moindres carrés ont trois origines :

- la spécification incomplète ou erronée du modèle

- l'usage des petits échantillons
- la corrélation des erreurs

Nous allons examiner successivement ces trois points, puis les méthodes d'estimation alternatives.

4.1. Spécification incomplète ou erronée du modèle

L'hypothèse implicite à toute étude économétrique est que le modèle que l'on estime est vrai et, très souvent, l'économiste s'en remet aux résultats même de l'étude économétrique pour conforter cette hypothèse. Or, dans le cas présent, il peut être imprudent d'induire l'existence de retards échelonnés de la seule significativité du coefficient du terme autoregressif car les modèles autoregressifs ne correspondent pas nécessairement à une distribution de retards. Un exemple emprunté à GRILICHES [1967] le montre aisément.

a) Distribution géométrique en autocorrélation des erreurs ?

Supposons que le modèle vrai soit une relation instantanée entre y_t et x_t ; mais que les résidus u_t suivent un processus d'autocorrélation du premier ordre :

$$(1) \begin{cases} y_t = a x_t + u_t \\ u_t = \rho u_{t-1} + \varepsilon_t \end{cases}$$

L'introduction d'une variable autoregressive donnera généralement des coefficients λ significatif et des résidus ε_t moins fortement corrélés :

$$(2) y_t = a x_t + \lambda y_{t-1} + \varepsilon_t$$

puisque le modèle vrai peut être écrit sous la forme :

$$(1') y_t = a x_t + \rho y_{t-1} - a \rho x_{t-1} + \varepsilon_t$$

Le modèle (2) peut ainsi correspondre soit à une distribution de retards géométrique, soit à une relation simultanée incomplètement spécifiée où les erreurs sont liées. Toutefois, dans le cas présent, l'introduction de la variable x_{t-1} permettra de choisir entre les deux hypothèses. Si en effet le coefficient de x_{t-1} est négatif et voisin du produit des coefficients de x_t et y_{t-1} on pourra considérer que l'hypothèse (1) est vraie et que (2) doit être rejetée. On remarquera que l'équation (1') est analogue à la forme autoregressive d'une distribution du premier ordre :

$$(3) \quad y_t = a_0 x_t + a_1 x_{t-1} + \lambda y_{t-1} + \varepsilon_t$$

Si en outre l'hypothèse (1) est vérifiée, on doit avoir :

$$a_0 \cdot \lambda = - a_1$$

et le délai moyen calculé $\theta = + \frac{a_1}{a_0 + a_1} + \frac{\lambda}{1-\lambda}$ est nul. Ainsi, si le modèle vrai est le modèle (1), une distribution géométrique (2) donnera une estimation biaisée du vrai modèle, mais une distribution du second ordre (3) montrera que l'ajustement est instantané. Il est souvent possible en pratique, en assouplissant les hypothèses relatives à la forme des retards échelonnés, d'approcher d'assez près la réalité économique. Dans le cas présent l'erreur d'interprétation résultant de l'estimation du modèle (2) peut être corrigée en prenant une distribution de retards plus souple (3) (le lecteur pourra consulter l'exemple empirique traité par GRILICHES page 34 de l'article cité). Cet exemple met également en lumière les biais résultant d'une spécification incomplète d'un modèle à retards échelonnés. Si la distribution réelle est en effet une distribution du premier ou du second ordre, l'estimation d'une distribution incomplète (par exemple géométrique), conduira à une estimation biaisée des délais moyens et des coefficients à long terme.

b) Spécification incomplète du modèle

Supposons par exemple que la distribution de retards soit une distribution du second ordre :

$$y_t = a x_t + b y_{t-1} + c y_{t-2} + u_t$$

et que l'on estime le modèle incomplet correspondant à la distribution géométrique :

$$y_t = a' x_t + b' y_{t-1} + u'_t$$

On obtient en supposant que x_t est stationnaire et sans corrélation :

$$\text{plim}(b') = \frac{b}{1-c}$$

Le délai moyen réel est égal à $\theta = \frac{b+2c}{1-b-c}$ puisque le modèle est du second ordre, alors que l'estimation du modèle incomplet donnera :

$$\text{plim } \theta' = \frac{\frac{b}{1-c}}{1 - \left(\frac{b}{1-c}\right)} = \frac{b}{1-b-c} > \theta$$

qui surestimera (c étant négatif) le délai moyen réel.

4.2. Les biais introduits par l'estimation sur petits échantillons

Si les erreurs du modèle autoregressif constituent un processus aléatoire sous corrélation temporelle, on montre* que les estimateurs des moindres carrés ordinaires convergent en probabilité vers les valeurs vraies des paramètres, quand le nombre d'observations augmente indéfiniment. Si elles sont en outre distribuées selon une loi normale, les estimateurs des moindres carrés sont asymptotiquement efficaces (leur variance est minimale parmi les estimateurs sans biais). On peut

./.

* cf. E. MALINVAUD [1969] chapitre 14 et JOHNSTON [1972] p. 304.

donc, comme nous l'avons fait dans la section précédente, appliquer les moindres carrés ordinaires comme si la variable endogène retardée était une variable exogène, mais la procédure n'est plus justifiée qu'asymptotiquement. En particulier, l'estimation sur des petits échantillons conduit à des erreurs systématiques.

Par exemple, pour le modèle simple autoregressif :

$$y_t = \lambda y_{t-1} + \varepsilon_t$$

HURWICZ [1950] a montré l'existence d'une erreur systématique qui, en première approximation est telle que :

$$E(\hat{b}) \simeq b \left(1 - \frac{2}{N}\right)$$

N étant le nombre d'observations

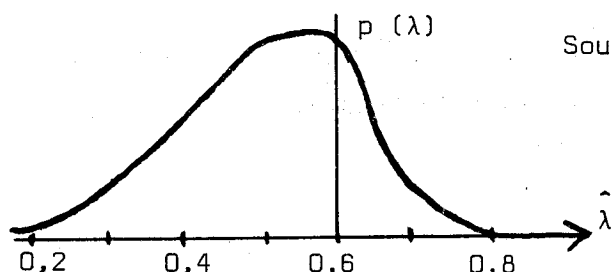
Pour une vingtaine d'observations (estimations sur série chronologique annuelle), le biais serait de l'ordre de 10 % de la valeur vraie.

Les modèles plus compliqués avec variable exogène ont été étudiés par les méthodes de "Monte Carlo", notamment par E. MALINVAUD [1961 - b -]. Dans ce dernier article, le modèle étudié comporte une variable endogène et un terme constant :

$$y_t = a x_t + \lambda y_{t-1} + d + \varepsilon_t$$

Des échantillons d'une vingtaine d'observations conduisent à des estimations $\hat{\lambda}$ qui sous-estiment dans trois cas sur quatre la valeur vraie du terme autoregressif (graphique [6]), le biais étant de l'ordre de (- 0,08) pour $\lambda = 0,6$.

GRAPHIQUE (6) - DENSITE DE DISTRIBUTION DE LA VALEUR ESTIME DU TERME AUTOREGRESSIF λ POUR 20 OBSERVATIONS



Source MALINVAUD [1961-b] ou [1969] page 579

La formule usuelle du calcul de l'écart type de λ donne en moyenne une estimation correcte, mais l'intervalle de confiance est décalé vers le bas puisque la valeur vraie est sous-estimée.

4.3. Corrélation temporelle des erreurs

Nous avons vu précédemment que les propriétés des moindres carrés ordinaires n'étaient valables qu'asymptotiquement pour un modèle autoregressif. Si, en outre les erreurs du modèle autoregressif sont corrélées, l'estimation par les moindres carrés perd à peu près toutes ses propriétés :

- l'estimateur est asymptotiquement biaisé
- il en est de même des estimateurs de la variance
- le coefficient de Durbin et Watson est biaisé et asymptotiquement peu efficace.

En particulier, les résidus de l'estimation par les moindres carrés sont en général moins corrélés que les erreurs du modèle autoregressif, si bien qu'il n'est pas possible d'apprécier correctement la liaison des erreurs à l'aide du test de DURBIN et WATSON.

Dans le cas d'une autocorrélation du premier ordre :

$$u_t = \rho u_{t-1} + \varepsilon_t$$

on peut appliquer le test de DURBIN [1970] qui repose sur des propriétés asymptotiques. Si D est la valeur du coefficient de DURBIN et WATSON, on obtient une estimation de ρ par $\hat{\rho} = 1 - \frac{D}{2}$

Le principe du test s'appuie sur le fait que la variable :

$$h = \frac{\hat{\rho}}{\sqrt{\frac{1}{N} - \text{Var}(\hat{\lambda})}}$$

Soit asymptotiquement une loi normale centrée réduite si $\rho=0$. L'hypothèse $\rho=0$ peut être par exemple rejetée au seuil de probabilité habituel si $h > 1,64$.

Les biais résultant de la corrélation temporelle des erreurs peuvent être importants, notamment pour le coefficient estimé du terme autoregressif. Considérons par exemple, d'un modèle autoregressif sans variable exogène

$$y_t = \lambda y_{t-1} + d + \epsilon_t$$

avec
$$\epsilon_t = \rho \epsilon_{t-1} + u_t$$

Lorsque le nombre d'observations augmente indéfiniment, le coefficient $\hat{\lambda}$ estimé par les moindres carrés tend vers

$$\text{plim } \hat{\lambda} = \frac{\lambda - \rho}{1 + \lambda\rho}$$

On voit que selon les valeurs du coefficient ρ , l'estimation de λ peut s'écarter de façon sensible de la valeur vraie λ et prendre notamment n'importe quelle valeur comprise entre 0 et 1. En particulier, pour $\rho=1$, $\text{plim } (\hat{\lambda})$ est égal à un quelle que soit la valeur vraie de λ . La relation précédente montre en outre que l'estimation surestime la valeur vraie lorsque la liaison des erreurs est positive et la sous-estime lorsqu'elle est négative.

L'introduction d'une variable exogène (x_t) :

$$y_t = a x_t + \lambda y_{t-1} + d + \epsilon_t$$

améliore cependant les qualités de l'estimation par les moindres carrés, notamment lorsque la variabilité de celle-ci excède celle des erreurs (cf. E. MALINVAUD [1969] chap. 14). Le biais asymptotique dépend en effet principalement de la grandeur $\alpha = \frac{\sigma_\epsilon^2}{a^2 \sigma_x^2}$, où σ_ϵ^2

est la variance des erreurs, σ_x^2 celle de la variable exogène x_t . Avec par exemple $\lambda = 0,6$ et $\rho = 0,5$, le biais est égal à 0,25 en l'absence de variable exogène, il tombe à 0,20 si la variable exogène a une faible variabilité ($\alpha=1$) et 0,08 si elle a au contraire une forte variabilité ($\alpha=0,25$).

4.4. Méthodes d'estimation alternatives

Le choix d'une méthode d'estimation dépend des hypothèses que l'on peut faire a priori sur les erreurs mais aussi de la taille de l'échantillon sur lequel on travaille, car les méthodes généralement proposées dans la littérature et que nous allons évoquer rapidement, reposent toutes sur des propriétés asymptotiques et il arrive fréquemment qu'elles ne soient pas plus performantes que les moindres carrés sur les échantillons habituellement utilisés en macroéconomie (par exemple une vingtaine d'observations pour les séries chronologiques annuelles). Les méthodes d'estimations dépendent des hypothèses que l'on peut faire a priori sur les erreurs des modèles à retards échelonnés ou de la forme autoregressive qui lui est associée. Nous noterons (ϵ_t) les erreurs du modèle à retards échelonnés:

$$(1) \quad y_t = \frac{P(L)}{1 - W(L)} x_t + \epsilon_t$$

et u_t les erreurs de la forme autoregressive :

$$(2) \quad y_t = W(L) y_t + P(L) x_t + u_t$$

Ces erreurs sont liées par la relation :

$$(3) \quad u_t = [1 - W(L)] \epsilon_t$$

Nous examinerons tout d'abord la méthode la plus générale, puis les méthodes associées à des hypothèses particulières sur les erreurs (ϵ_t) ou (u_t) .

a) La méthode la plus générale consiste à utiliser, comme variables instrumentales, les variables endogènes retardées. Elle ne nécessite pas en effet d'hypothèses particulières sur u_t , si ce n'est que leur espérance mathématique est nulle. Considérons par exemple la distribution géométrique :

$$(2) \quad y_t = \lambda y_{t-1} + a x_t + u_t$$

L'existence d'un biais asymptotique dans l'ajustement de la relation (2) par les moindres carrés ordinaires tient au fait que y_{t-1} et u_t ne sont pas indépendants lorsqu'il existe une liaison entre u_t et u_{t-1} . Si l'on remplace en revanche y_{t-1} par x_{t-1} dans les équations normales (x_t et x_{t-1} étant les variables instrumentales), les estimations obtenues sont sans biais asymptotique (cf. MALINVAUD [1969] chap. 14).

Dans le cas plus général d'une distribution d'ordre n , on utiliserait la variable x_t et n valeurs retardées x_{t-i} comme variables instrumentales.

Toutefois, si les erreurs sont indépendantes, les coefficients estimés par cette méthode ont des écarts-types supérieurs à ceux des moindres carrés ordinaires*. En pratique, lorsque le biais asymptotique est faible, la perte de précision a pour conséquence que les moindres carrés ordinaires sont souvent plus performants que la méthode des variables instrumentales. Ce résultat a été notamment confirmé par l'étude d'échantillons artificiels comportant une vingtaine d'observations (E. MALINVAUD [1961-b] et l'écart entre les variables instrumentales et les moindres carrés ordinaires est encore plus important sur les séries chronologiques annuelles car x_t et x_{t-1} sont souvent corrélés. Pour l'accélérateur flexible sur série chronologique annuelle, l'estimation par les variables instrumentales conduisait par exemple à des coefficients $\hat{\alpha}$ et $\hat{\lambda}$ qui n'étaient plus significativement différents de zéro (cf P.A. MUET, P. ZAGAME [1976]). Nous verrons plus loin, en comparant différentes méthodes d'estimations d'une distribution du second ordre pour l'accélérateur flexible, que la précision des variables instrumentales est également plus faible que celle des moindres carrés ordinaires, mais que les coefficients restent cependant significativement différents de zéro.

./.

* Les écarts-types calculés par les programmes de régression standards sont donc supérieurs à ceux des moindres carrés ordinaires.

b) Hypothèses-relatives aux erreurs du modèle à retards échelonné

Si nous supposons que les ϵ_t constituent un processus purement aléatoire, les erreurs de la forme autoregressive (u_t) forment alors un processus de moyenne mobile dont le corrélogramme dépend des paramètres $W(L)$ du modèle autoregressif. Par exemple, dans l'hypothèse d'une distribution géométrique le coefficient de corrélation entre u_t et u_{t-1} (ou coefficient d'autocorrélation du premier ordre) est négatif et compris entre 0 et 0,5 (il est égal à $r_1 = -\frac{\lambda}{1 + \lambda^2}$), tandis que les autres coefficients du corrélogramme sont nuls. Une méthode évidente d'estimation consiste à estimer directement les coefficients de retards sur la forme à retards échelonnés du modèle, par les méthodes applicables aux modèles non linéaires (maximum de vraisemblance, qui équivaut dans le cas présent à minimiser la somme des carrés des résidus). DHRYMES [1971] suggère ainsi une estimation directe qui consiste à minimiser en deux temps la variance résiduelle.

Considérons par exemple la distribution géométrique :

$$(1) y_t = \frac{a}{1 - \lambda L} x_t + \epsilon_t \quad 1 \leq t \leq N,$$

sous l'hypothèse de normalité des erreurs : $\epsilon \sim N(0, \sigma^2)$.

Ce modèle peut s'écrire sous la forme :

$$y_t = \sum_{i=0}^{t-1} \lambda^i x_{t-i} + \lambda^t \sum_{i=0}^{\infty} \lambda^i x_{-i}$$

la somme $a_0 = \sum_{i=0}^{\infty} \lambda^i x_{-i}$ ne dépend pas de la période d'estimation, on est donc ramené à l'estimation de a_0 , λ , et a définis par :

$$(3) y_t = a_0 \lambda^t + a \sum_{i=0}^{t-1} \lambda^i x_{t-i} + \epsilon_t \quad 1 \leq t \leq N$$

./.

L'estimation du maximum de vraisemblance s'obtient facilement en minimisant la variance résiduelle. On peut déterminer ce minimum en deux temps. On se fixe une valeur de λ et on calcule $(\hat{a}, \hat{a}_0, \hat{\sigma}^2(\lambda))$ en appliquant les moindres carrés ordinaires à la relation :

$$(3') \quad y_t = a_0 \lambda^t + a z_t + \epsilon_t$$

$$\text{avec} \quad z_t = \sum_{i=0}^{t-1} \lambda^i x_{t-i}$$

On cherche ensuite le minimum de $\sigma^2(\lambda)$ en effectuant l'opération pour diverses valeurs de λ .

En pratique, cette méthode a peu d'intérêt car elle devient rapidement assez lourde (notamment pour une distribution du second ordre) et ses propriétés reposent comme toujours sur la validité des hypothèses (le modèle est vrai, les ϵ_t sont sans corrélation et distribués selon une loi normale). Or l'économiste utilise une distribution géométrique ou du second ordre non parce qu'il considère que c'est la forme exacte des retards mais comme approximation simple de la distribution exacte : l'estimer sur la forme à retards échelonnés lui fait perdre tout son intérêt empirique pour des hypothèses qui ne sont guère plus assurées que celles d'indépendance des erreurs de la forme autoregressive.

Une méthode d'estimation a été proposée par KOYCK [1954] et KLEIN [1958]. Elle se situe à mi-chemin entre les moindres carrés généralisés et les moindres carrés ordinaires. KOYCK et KLEIN tiennent compte du fait que la variance de u_t dépend de λ , mais négligent le fait que u_t et u_{t-1} sont corrélés. La matrice des variances-covariances de u_t est en effet :

$$V(u_t) = \sigma^2 \begin{bmatrix} (1+\lambda^2) & -\lambda & 0 & \dots & 0 \\ -\lambda & (1+\lambda^2) & -\lambda & \dots & 0 \\ 0 & 0 & 0 & \dots & (1+\lambda^2) \end{bmatrix}$$

et la méthode de KOYCK et KLEIN ne tient pas compte des termes non diagonaux de la matrice et présente de ce fait un intérêt limité.

Si l'on suppose plus généralement que les erreurs du modèle à retards échelonnés suivent un processus d'autocorrélation du premier ordre :

$$\varepsilon_t = \rho \varepsilon_{t-1} + \eta_t$$

diverses méthodes permettent d'estimer le modèle dans cette hypothèse (maximum de vraisemblance, estimation en deux temps combinant les variables instrumentales et les moindres carrés généralisés). Elles ont été particulièrement étudiées par DHRYMES [1971] d'un point de vue théorique (propriétés asymptotiques, efficacité) et empirique (méthode de Monte-Carlo, page 355 et suivantes). Nous reviendrons sur ce dernier point dans le paragraphe suivant. Mentionnons pour terminer l'application de l'analyse spectrale suggérée initialement par HANNAN [1967] et développée par de nombreux auteurs. Cette méthode applicable à des erreurs (ε_t) suivant des processus stochastiques beaucoup plus généraux que les deux hypothèses précédentes ne semble pas avoir en revanche un grand intérêt pour les études empiriques.

c) Hypothèses relatives aux erreurs de la forme autoregressive

Pour avoir une idée de la validité de l'hypothèse implicite à l'usage des moindres carrés ordinaires sur la forme autoregressive il nous faut étudier l'autocorrélation des résidus (u_t) de la forme autoregressive, pour diverses hypothèses relatives aux (ε_t). Supposons comme précédemment que ceux-ci présentent une autocorrélation du premier ordre :

$$\varepsilon_t = \rho \varepsilon_{t-1} + \eta_t$$

Les erreurs de la forme autoregressive suivent alors un processus de la forme :

$$u_t = (1-\lambda L) (1-\rho L)^{-1} \eta_t,$$

dont le coefficient d'autocorrélation d'ordre k est :

$$r_k = \rho^{k-1} \frac{\rho(1+\lambda^2) - \lambda(1+\rho^2)}{(1+\lambda^2) - 2\lambda\rho}$$

Le graphique (7) indique les variations du coefficient d'autocorrélation du premier ordre (r_1) des erreurs de la forme autoregressive. Si les erreurs du modèle à retards échelonnés sont indépendantes, celles de la forme autoregressive présentent une autocorrélation négative. Celle-ci est encore accentuée si les ε_t présentent déjà une autocorrélation négative. Lorsqu'en revanche les ε_t présentent déjà une autocorrélation positive, l'autocorrélation des erreurs de la forme autoregressive est faible, et l'usage des moindres carrés ordinaires peut être largement justifié (les u_t sont évidemment indépendants lorsque $\rho=\lambda$). En pratique on pourra donc souvent appliquer les moindres carrés car l'autocorrélation des erreurs de la forme autoregressive a des chances d'être faible et de plus, le biais asymptotique sera d'autant plus faible que les variations de la variable endogène (x_t) seront importantes. En outre, l'étude par les méthodes de Monte-Carlo précédemment mentionnée (DHRYMES) conforte assez largement cette conclusion : "les moindres carrés ordinaires sont un assez bon estimateur du modèle lorsque λ est fort et que la variable endogène x_t présente une autocorrélation positive. Cet aspect qui risque de décourager les approfondissements analytiques ultérieurs car la combinaison des deux éventualités semble assez fréquentes dans les travaux empiriques" (DRYMES [1971] p. 400)*

Au terme de cet examen rapide nous allons, en guise de conclusion, présenter une application empirique des méthodes qui constituent à la fois un complément et une alternative simples aux moindres carrés ordinaires :

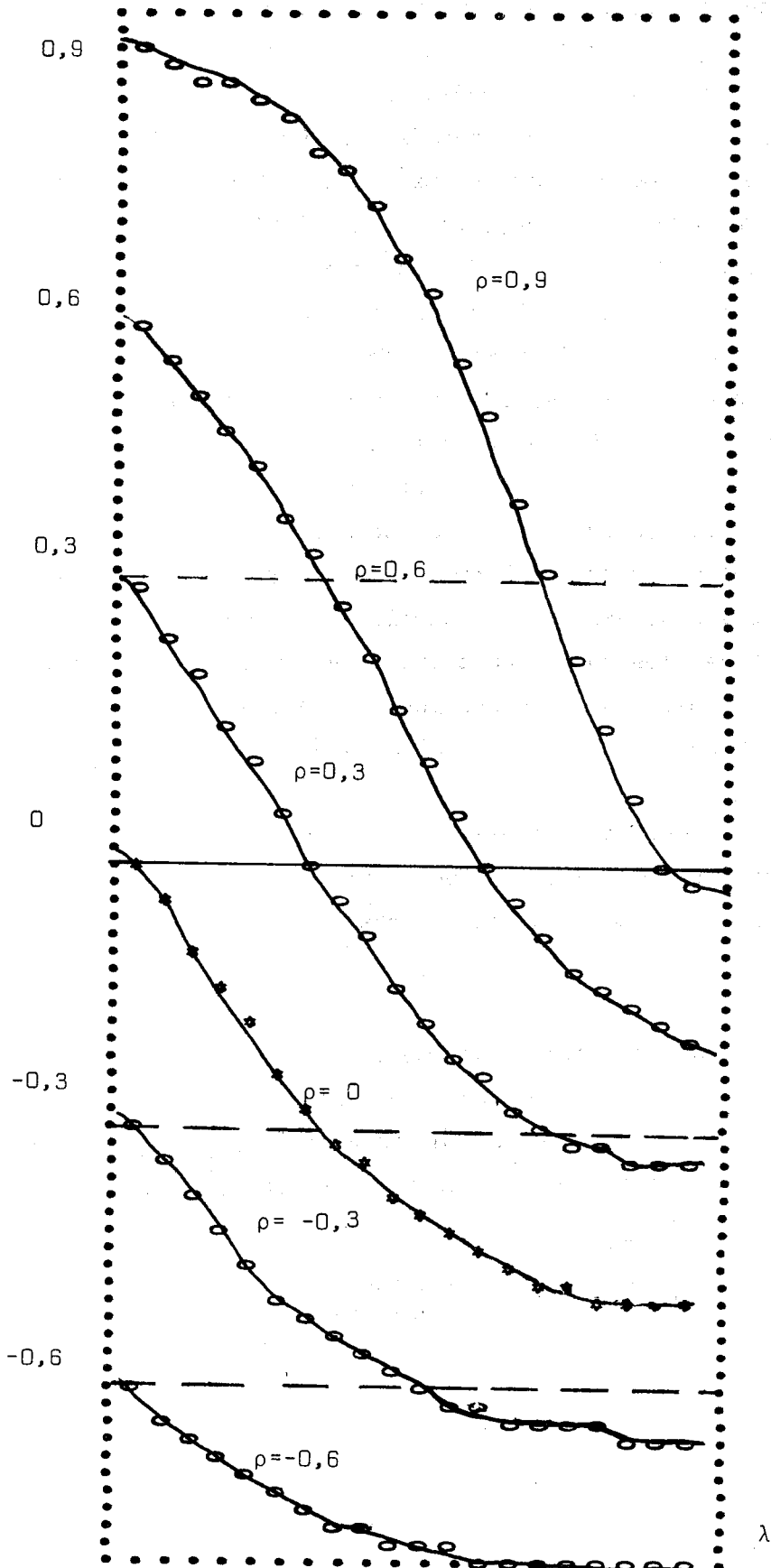
- les variables instrumentales (a)
- la prise en compte d'une autocorrélation du premier ordre des erreurs de la forme autoregressive, qui peut apparaître comme une bonne approximation d'hypothèses alternatives (b).

----- ./.

* Cette remarque ne manque pas d'intérêt au terme d'un long ouvrage consacré à l'étude analytique de méthodes d'estimations souvent fort "sophistiquées".

GRAPHIQUE (7) - VARIATIONS DU COEFFICIENT D'AUTOCORRELATION
DU PREMIER ORDRE DES ERREURS DE LA FORME AUTOREGRESSIVE AVEC

λ ET ρ



4.5. Un exemple : distribution du second ordre de l'accélérateur flexible

L'estimation par les moindres carrés ordinaires d'une distribution du second ordre de l'accélérateur flexible fait apparaître sur la période 1953-1974 une autocorrélation négative des erreurs. Le simple examen de la statistique de DURBIN et WATSON (DW = 2,7) le suggère mais on peut, en appliquant le test de DURBIN, montrer que l'hypothèse d'indépendance des erreurs peut être rejetée.

Pour éliminer le biais asymptotique, deux méthodes simples sont envisageables :

- l'estimation par les variables instrumentales (x_{t-1} , x_{t-2})
- l'estimation des coefficients et de l'autocorrélation (ρ) par la méthode de COCHRANE-ORCUTT.

Le tableau (8) ci-après présente l'estimation des coefficients a, b et c de l'accélérateur flexible :

$$\frac{I}{K} = a Q^0 + b \left(\frac{I}{K}\right)_{-1} + c \left(\frac{I}{K}\right)_{-2} + d,$$

par les trois méthodes alternatives (moindres carrés ordinaires, variables instrumentales, COCHRANE-ORCUTT). Les estimations par les moindres carrés ordinaires et par les variables instrumentales ont été réalisées sur la période 1953-1974 et comportent 22 observations, l'estimation par la méthode de COCHRANE-ORCUTT l'a été sur la période 1954-1974 (21 observations). Les (T) de Student des coefficients estimés ont été indiqués entre parenthèses.

TABLEAU (8) - COMPARAISON DE TROIS METHODES D'ESTIMATION
D'UNE DISTRIBUTION DU SECOND ORDRE

	\hat{a}	\hat{b}	\hat{c}	R^2	DW	$\Sigma e^2 \cdot 10^{-4}$
MCO (1953-1974)	0,246 (5,7)	0,896 (8,9)	-0,202 (2,0)	0,93	2,7	1,21
VI (1953-1974)	0,233 (5,1)	0,994 (7,1)	-0,269 (1,5)	-	2,8	1,28
CORC (1954-1974) $\hat{\rho} = -0,19$ (2,5)	0,259 (4,6)	0,997 (10,1)	-0,298 (3,4)	0,92	2,2	0,95

Les trois méthodes donnent des coefficients très voisins et non significativement différents. En particulier, des coefficients estimés par la méthode de COCHRANE-ORCUTT et par la méthode des variables instrumentales sont pratiquement identiques. Leur comparaison avec les estimations des moindres carrés peut donner une idée du biais asymptotique : \hat{b} serait ainsi sous-estimé de 0,10 et la valeur absolue de \hat{c} le serait un peu moins (0,07 à 0,10). Pour les deux coefficients, la différence est de l'ordre de l'écart-type et n'est donc pas significative. Ainsi, en dépit d'une corrélation évidente des erreurs du modèle autoregressif ($\hat{\rho}$ est significativement différent de zéro), le biais reste inférieur à la précision de l'estimation. On remarquera enfin que la précision des variables instrumentales est plus faible que celle des moindres carrés et que la méthode de COCHRANE-ORCUTT donne l'estimation la plus précise pour les coefficients \hat{b} et \hat{c} .

5 - LES DISTRIBUTIONS FINIES ET LEUR APPROXIMATION POLYNOMIALE

Les distributions finies faisant intervenir un nombre élevé de retards échelonnés sont souvent difficiles à estimer en raison notamment de la corrélation des valeurs retardées des variables explicatives. Aussi doit on généralement imposer des contraintes aux coefficients de la distribution de retards. Entre une distribution entièrement spécifiée comme la distribution géométrique et l'estimation directe des $(s+1)$ coefficients w_θ d'un modèle à retards échelonnés :

$$(1) \quad y_t = w_0 x_t + w_1 x_{t-1} + w_2 x_{t-2} + \dots + w_s x_{t-s} + \varepsilon_t$$

On peut choisir de nombreuses fonctions représentatives des coefficients w_θ , comportant un nombre de paramètres inférieurs à $(s+1)$. La méthode suggérée par ALMON [1965], qui consiste à ajuster un polynôme de degré r sur les coefficients w_θ des variables retardées, présente l'avantage d'être peu contraignante quant à la forme de la distribution des retards et de se prêter à une estimation économétrique relativement facile. Sa souplesse explique le succès rencontré par cette méthode dans les études réalisées sur données trimestrielles.

5.1. La méthode d'ALMON

Supposons que les $(s+1)$ coefficients w_θ soient distribués selon un polynôme de degré r de la variable θ :

$$(1) \quad w_\theta = \beta_0 + \beta_1 \theta + \dots + \beta_r \theta^r \quad \theta = 0, \dots, s$$

La méthode d'ALMON consiste à remplacer l'estimation des $(s+1)$ coefficients w_θ par celle des $(r+1)$ coefficients β_i du polynôme. La relation (1) s'écrit sous forme matricielle :

$$(2) \quad \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_s \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 1^0 & 1^1 & 1^2 & \dots & 1^r \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s^0 & s^1 & s^2 & \dots & s^r \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_r \end{pmatrix} \quad r < s$$

et l'équation à estimer (équation 1) devient :

$$(3) \quad y_t = \beta_0 (x_t + x_{t-1} + \dots + x_{t-s}) + \beta_1 (x_{t-1} + 2x_{t-2} + \dots + sx_{t-s}) + \dots + \beta_r (x_{t-1} + 2^r x_{t-2} + \dots + s^r x_{t-s}) + \varepsilon_t$$

L'estimation des coefficients β_i s'obtient en appliquant les moindres carrés ordinaires à la relation précédente :

$$(3) \quad y_t = \beta_0 z_{0t} + \beta_1 z_{1t} + \dots + \beta_r z_{rt} + \varepsilon_t$$

où $z_{it} = (x_{t-1} + 2^i x_{t-2} + \dots + s^i x_{t-i})$

Le degré du polynôme étant peu élevé (troisième ou quatrième degré) le modèle comporte peu de variables explicatives.

Les coefficients \hat{w}_θ se déduisent des coefficients $\hat{\beta}_i$ estimés par la relation (2) :

$$\hat{w} = S \hat{\beta}$$

où S est la matrice précédemment définie.

Cette méthode d'estimation des coefficients du polynôme présente l'avantage de permettre de tester directement les hypothèses retenues pour le degré du polynôme. On peut en effet choisir entre le degré (r) et (r-1) en appliquant au coefficient $\hat{\beta}_r$ les tests classiques d'analyse de la variance. Toutefois, l'estimation des coefficients $\beta_0 \dots \beta_r$ peut s'avérer délicate car les variables z_i sont souvent fortement corrélées et la matrice S est en outre, et pour la même raison, difficilement inversible. C'est la raison pour laquelle on a parfois recours à une spécification différente du polynôme, initialement suggérée par ALMON, et qui utilise l'interpolation de LAGRANGE. Le polynôme s'écrit alors :

$$f(\theta) = \phi(\theta, \theta_0) f(\theta_0) + \dots + \phi(\theta, \theta_r) f(\theta_r)$$

où $\phi(\theta, \theta_i)$ est le polynôme d'interpolation de LAGRANGE. Les calculs sont analogues mais on peut choisir les θ_i de façon à ce que la matrice S soit plus facilement inversible. Si le calcul numérique en est facilité, le résultat de l'estimation, c'est-à-dire les coefficients \hat{w}_θ estimés sont identiques dans les deux méthodes et ne dépendent pas de la paramétrisation du polynôme (voir par exemple DHRYMES [1971] page 230 et suivantes).

En pratique, la méthode d'ALMON suppose donc que soit fixé le degré (r) du polynôme, la longueur des délais (s) et les éventuelles contraintes aux limites de la distribution. Le choix du degré et la longueur des délais s'opèrent en comparant les estimations correspondant à différentes valeurs de r et s. En général, on retient un polynôme de faible degré (3 ou 4) et on choisit s en retenant la valeur qui conduit à la plus faible variance résiduelle, ou au plus grand R^2 puisque lorsque le degré du polynôme est fixé, le nombre de variables explicatives est également fixé.

5.2. Un exemple : l'accélérateur estimé sur série trimestrielle

Le modèle estimé est le suivant* :

$$\left(\frac{I}{Q}\right)_t = \sum_{\theta=0}^s w_{\theta} Q_{t-\theta} + c t + d$$

Q est la production Q sont taux de croissance, I l'investissement et t un trend temporel représentant l'augmentation tendancielle du taux d'investissement non expliquée par le modèle d'accélération. La distribution de retards est supposée être un polynôme du troisième degré, sans contrainte aux extrémités de l'intervalle $[0,5]$. Selon l'usage habituel, s est déterminé en cherchant le minimum de la variance résiduelle (ou le maximum du R^2), pour différentes valeurs de s . Le graphique (7) représente la valeur du R^2 et la forme des distributions correspondant à différents échelonnements des retards ($5 \leq s \leq 10$ trimestres). Le maximum correspond à $s = 9$ trimestres mais la valeur des coefficients w_{θ} n'est plus significativement différente de zéro au-delà de 7 trimestres et en outre, les distributions de retards obtenues pour des valeurs de s comprises entre 6 et 10 trimestres sont pratiquement confondues, aussi peut-on retenir la distribution correspondant à $s = 7$ trimestres dont les coefficients estimés sont tous positifs et significativement différents de zéro.

Cette méthode de recherche de la "meilleure" structure de retards, par "balayage" sur les valeurs de s n'est pas très satisfaisante car il n'existe pas de tests permettant de choisir entre deux modèles différents (par exemple entre $s = 6$ et $s = 7$), et elle conduit en outre à surestimer la valeur vraie de s (cf paragraphe 54). En revanche, les hypothèses relatives au degré du polynôme peuvent être testées par l'analyse de la variance ou dans le cas d'une estimation directe des coefficients du polynôme par la valeur du (T) de Student des coefficients.

./.

* Pour plus de détails on pourra se reporter à P.A. MUET, P. ZAGAME [1976].

Dans l'estimation considérée, le logiciel utilisé (TSP) estime directement les coefficients β_i du polynôme :

$$w_\theta = \beta_1 + \beta_2 \left(\frac{\theta+1}{s+2}\right)^1 + \dots + \beta_{r+1} \left(\frac{\theta+1}{s+2}\right)^r$$

Le tableau (9) indique la valeur des coefficients estimés ($\hat{\beta}_i$) ainsi que leur T de Student. La précision dans l'estimation des coefficients (β_i) est très médiocre car, comme nous l'avons indiqué, les variables z_{it} sur lesquelles porte l'estimation sont assez fortement colinéaires. On remarquera que le coefficient β_1 n'est jamais significativement différent de zéro c'est-à-dire que l'hypothèse que w_{-1} est nul ne peut jamais être rejetée. On pourrait donc contraindre la distribution à s'annuler au point $\theta=-1$. On pourrait tirer la même conclusion pour l'autre extrémité de la distribution car la somme des 4 coefficients qui est égale à w_{s+1} n'est pas non plus significativement différente de zéro. La valeur du T de Student du coefficient β_4 montre que le polynôme retenu pourrait-être du second degré lorsque $s \leq 8$ mais qu'au-delà, le troisième degré est significatif, ce que suggère bien d'ailleurs la forme des distributions de retards qui est assimilable à des paraboles lorsque s est inférieur ou égal à 8, mais présente ensuite une inflexion (graphique [8]).

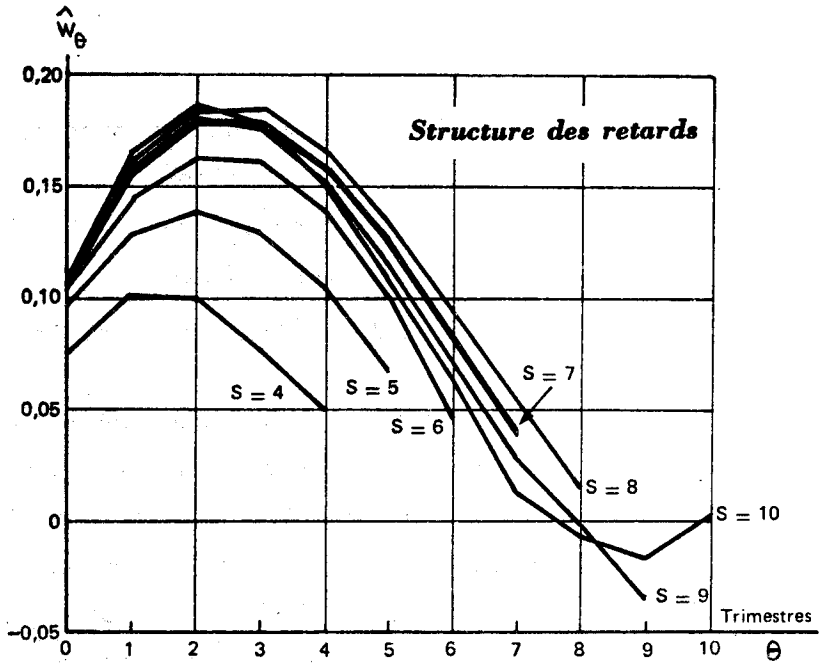
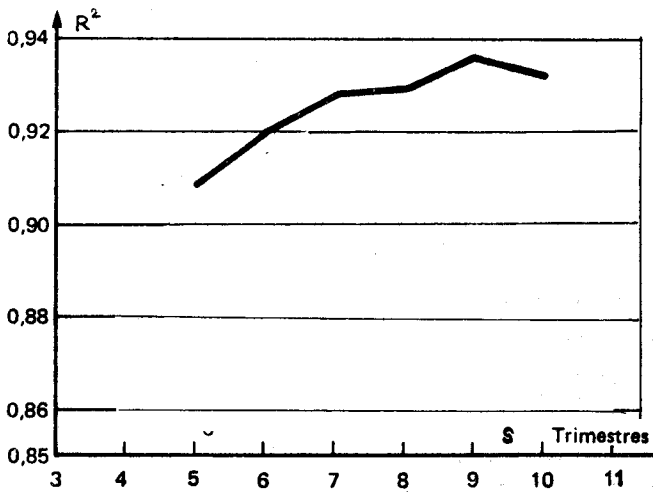
TABLEAU (9) - VALEUR DES $\hat{\beta}_i$ ET DE LEUR T DE STUDENT
=====

S	6	7	8	9	10
$\hat{\beta}_1$	0,04 (0,8)	0,02 (0,4)	0,02 (0,4)	0,02 (0,4)	0,00 (0,01)
$\hat{\beta}_2$	0,60 (1,3)	0,95 (2,3)	1,09 (2,8)	1,23 (3,3)	1,55 (4,1)
$\hat{\beta}_3$	-0,80 (0,7)	-1,62 (1,7)	-2,05 (2,4)	-2,67 (3,1)	-3,85 (4,4)
$\hat{\beta}_4$	0,13 (0,2)	0,65 (1,1)	0,93 (1,7)	1,38 (2,6)	2,36 (4,2)

GRAPHIQUE (8) - FORME DES DISTRIBUTIONS ET
VALEUR DU R^2 POUR DIFFERENTES VALEURS DE LA LONGUEUR
MAXIMALE DES RETARDS (s)

Source - P.A.MUET, P.ZAGAME [1976]

Valeurs du coefficient de corrélation multiple R^2



5.3. Méthode d'ALMON ou moindres carrés ordinaires : Test de l'hypothèse d'une distribution polynômiale

L'estimation d'une distribution polynômiale équivaut à estimer les paramètres d'un modèle linéaire

$$y_t = \sum_{\theta=0}^s w_{\theta} x_{t-\theta} + \epsilon_t$$

sujet à des contraintes linéaires représentées par la relation matricielle :

$$w = S \cdot \beta$$

Il est important en pratique de savoir si l'hypothèse selon laquelle les coefficients de retards w suivent une distribution polynômiale peut être retenue contre l'hypothèse que les coefficients w sont sans contraintes. Autrement dit, tester l'hypothèse H_0 : $w = S \beta$, contre l'hypothèse H_1 : w est sans contrainte.

L'analyse classique de la variance permet de réaliser ce test*. On peut le faire pour différentes valeurs de s , ou pour une valeur fixée a priori. Par exemple dans le modèle estimé précédemment, nous avons vu que l'on pouvait retenir $s = 7$ trimestres. L'estimation par la méthode d'ALMON donne alors une structure de retards dont les coefficients w_{θ} sont positifs et significativement différents de zéro et qui se situent par hypothèse sur un polynôme du troisième degré ($r=3$). Cette hypothèse est-elle acceptable ou doit-on au contraire la rejeter ? Considérons la période allant du 4ème trimestre 1964 au 4ème trimestre 1973 sur laquelle a porté l'estimation ($N=37$ observations), et désignons par Q_0 la variance résiduelle de l'estimation de la distribution polynômiale :

$$Q_0 = \epsilon' [I - X S (S' X' X S)^{-1} S X'] \epsilon ,$$

et par Q_1 la variance résiduelle de l'estimation directe du modèle à retards échelonnés par les moindres carrés ordinaires :

$$Q_1 = \epsilon' [I - X (X' X)^{-1} X'] \epsilon$$

./.

* Ce test est présenté dans DHRYMES [1971] pages 227-9. Un test plus général a été établi par G. GODFREY et D.S POSKITT [1975], il ne nécessite pas l'estimation du modèle contraint.

La première estimation comporte $N-(r+1) - 2$ degrés de libertés ($r+1$ variables explicatives pour la distribution polynômiale plus le terme constant, plus le trend) ; la seconde $N-(s+1)-2$ ($s+1$ retards échelonnés, plus le terme constant et le trend). Sous les hypothèses habituelles de normalités des erreurs :

$$\varepsilon \sim N(0, \sigma^2)$$

on a donc :

$$\text{ALMON} \quad \frac{Q_0}{\sigma^2} \sim \chi^2_{N-(r+1)-2}$$

$$\text{MCO} \quad \frac{Q_1}{\sigma^2} \sim \chi^2_{N-(s+1)-2}$$

et comme les matrices des formes quadratiques sont orthogonales :

$$\frac{Q_0 - Q_1}{\sigma^2} \sim \chi^2_{s-r}$$

Au seuil de confiance $(1-\alpha)$, l'hypothèse selon laquelle les coefficients de retards peuvent être représentés par une distribution polynômiale de degré r et de longueur s peut être rejetée si :

$$\frac{Q_0 - Q_1}{Q_1} \cdot \frac{N-r-3}{s-r} > F_{s-r, N-r-3}(\alpha)$$

Dans le cas examiné ($r=3, N=37, s=7$), et pour le seuil $\alpha=5\%$ on obtient :

$$\text{MCO} \quad Q_1 = 0,1771 \cdot 10^{-3}$$

$$\text{ALMON} \quad Q_0 = 0,2028 \cdot 10^{-3}$$

$$\text{et} \quad \frac{Q_0 - Q_1}{Q_1} \cdot \frac{31}{4} = 1,12 < F_{4,31}(0,05) = 2,7$$

L'hypothèse selon laquelle les retards échelonnés peuvent être représentés par une distribution polynômiale du troisième degré ne peut être rejetée par rapport à l'hypothèse alternative d'une distribution quelconque. Ce résultat apparaît intuitivement lorsqu'on examine la forme des retards échelonnés résultant de l'estimation directe des

w_θ par les moindres carrés ordinaires pour $s=7$. Les retards échelonnés obtenus ainsi sont très proches de la distribution polynômiale estimée comme le montre le graphique (9). Il en est de même délais moyens et des coefficients à long terme comme cela apparaît dans le tableau (10) ci-après.

TABLEAU (10) - DISTRIBUTION POLYNOMIALE ET DISTRIBUTION QUELCONQUE (MCO)
COMPARAISON DES AJUSTEMENTS POUR s=7 TRIMESTRES

	CLT*	Délai moyen (θ)*	R^2	D	Σe^2 . 10^3
Distribution Polynômiale (r=3)	1,02 (0,18)	3,00 (0,45)	0,928	1,5	0,2028
Distribution quelconque (MCO)	1,02 (0,18)	2,98 (0,21)	0,937	1,3	0,1771

* L'écart-type est indiqué entre parenthèses.

5.4. Les limites de la méthode d'ALMON

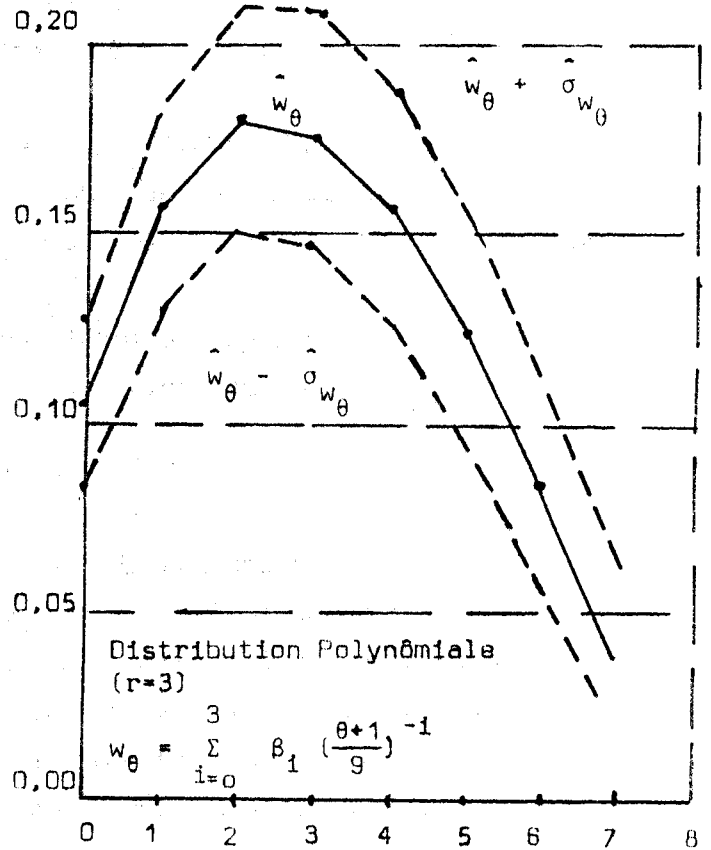
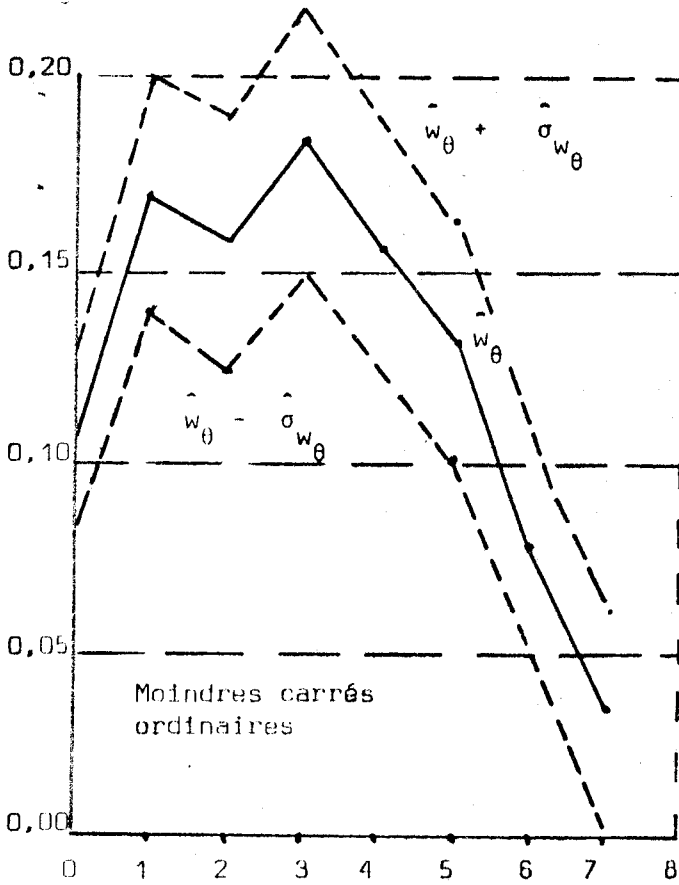
Lorsque la longueur des délais (s) et le degré du polynôme (r) sont connus, l'estimation des coefficients par la méthode d'ALMON possède toutes les propriétés de l'estimateur des moindres carrés ordinaires sous réserve des hypothèses habituelles concernant les erreurs ϵ_t . Elle est même plus efficace que l'estimation directe des coefficients de retards w_θ par l'application des moindres carrés au modèle :

$$y_t = \sum_{\theta=0}^s w_\theta x_{t-\theta} ,$$

puisque celle-ci ne tient pas compte de la contrainte qui lie les coefficients w_θ ($w=s \beta$).

GRAPHIQUE (9) - RETARDS ECHELONNES DE L'ACCELERATEUR FLEXIBLE :

DISTRIBUTION POLYNOMIALE ET DISTRIBUTION NON CONTRAINTE (MOINDRES
CARRÉS ORDINAIRES)



Mais cette hypothèse n'est pratiquement jamais réalisée car on ne connaît pas à priori la valeur vraie des paramètres s et r et la méthode qui consiste à choisir la "meilleure" structure de retards en faisant varier les deux paramètres conduit à des estimations biaisées des coefficients w_θ et de leur variance.

Les biais induits par une spécification erronée du polynôme ont été mis en évidence par SCHMIDT et WAUD [1973], puis étudiés par SCHMIDT et SICKLES [1975] et P.A. FROST [1975].

Supposons que le modèle vrai soit par exemple le modèle à retards échelonnés sans contrainte :

$$y_t = \sum_{\theta=0}^s w_\theta x_{t-\theta}$$

et que l'on estime une distribution polynômiale de degré $r < s$. Cette estimation des coefficients w_θ sera biaisée, ainsi que la matrice des variances-covariances $V(w)$. Il en sera de même si la distribution vraie est une distribution polynômiale de degré r^* et de longueur s^* et que l'on estime une distribution de degré $r \neq r^*$ et de longueur $s \neq s^*$. Malgré cela, la variance résiduelle de l'estimation de la distribution (r,s) peut être inférieure à celle de la distribution vraie (r^*, s^*) , si bien que la méthode de recherche de la distribution optimale par minimisation de la variance résiduelle a peu de chance de conduire à une estimation correcte de la distribution réelle.

Lorsqu'on cherche en effet à déterminer à la fois les coefficients w_θ de la distribution et les caractéristiques (r^*, s^*) du polynôme, ces dernières ne sont plus des variables certaines mais des variables aléatoires. Les variances calculées ignorent donc la variance résultant de l'estimation des paramètres s^* et r^* . P.A FROST [1975] a étudié par les méthodes de Monte Carlo les conséquences de cette méthode d'estimation pour une distribution polynômiale du troisième degré s'étendant sur 10 périodes. La conclusion est que la méthode d'estimation usuelle entraîne des biais assez faibles pour les coefficients de retards w_θ , mais importants pour la variance.

En outre, comme le montre le tableau (11) ci-après, la spécification retenue pour le polynôme varie considérablement : le polynôme du 6ème degré est choisi dans 37 % des cas tandis que la bonne spécification n'est choisie dans aucune des 100 estimations. Enfin, en moyenne, cette méthode d'estimation tend à surestimer la valeur du degré (74 % des cas) et la longueur des retards (57 % des cas).

TABLEAU (11) - DISTRIBUTION DU DEGRE (r) ET DE LA LONGUEUR DES RETARDS (s)
OBTENUES PAR MINIMISATION DE LA VARIANCE RESIDUELLE

Longueur des retards (s) \ Degré du polynôme (r)	r*											Total
	6	7	8	9	10	11	12	13	14	15	16	
2	0	1	0	3	2	1	1	0	1	0	1	10
r* = 3	0	1	2	1	0	0	3	2	0	5	2	16
4	0	3	2	1	2	0	0	1	1	4	5	19
5	1	1	2	1	0	2	4	1	2	2	2	18
6	4	1	7	4	4	1	3	4	2	2	5	37
Total	5	7	13	10	8	4	11	8	6	13	15	100

Source : P.A. FROST [1975]

5.5. Généralisations de la distribution polynômiale : coefficients variables et coefficients stochastiques

De nombreuses généralisations de la distribution polynômiale ont été proposées. L'une des plus utilisée dans les modèles économétriques

consiste à supposer que les coefficients w_θ de la distribution sont eux-mêmes fonctions d'une variable conjoncturelle par exemple $z_{t-\theta}$:

$$y_t = \sum_{\theta=0}^s w_\theta x_{t-\theta} + d$$

$$w_\theta = a_\theta z_{t-\theta} + b_\theta$$

Le modèle complet s'estime simplement en appliquant la méthode d'ALMON aux variables $(x_{t-\theta})$ et $(x_{t-\theta} \cdot z_{t-\theta})$. Cette méthode a été appliquée notamment aux délais qui séparent les commandes des livraisons de biens d'équipement par TINSLEY [1967] et MUET, ZAGAME [1976], ainsi qu'aux délais de réalisation de l'investissement, dans le cadre d'une fonction globale d'investissement, par ANDO, MODIGLIANI, RAASCHE, TURNOVSKY [1974] et de MENIL et YOHN [1978].

SCHMIDT [1974] a également proposé une combinaison de la distribution géométrique et polynômiale :

$$w_\theta = \lambda^\theta \sum_{i=0}^r \beta_i \theta^i \quad \theta \in [0, \infty [$$

Enfin l'hypothèse d'une distribution stochastique des coefficients (w_θ) a été avancée par SCHILLER [1973] et étudiée par MADDALA [1977] puis ULLAH et RAJ [1978]. La forme la plus générale consiste à supposer que les coefficients w_θ sont la somme d'une composante \bar{w}_θ polynômiale et d'une composante stochastique $u_{\theta t}$:

$$w_\theta = \bar{w}_\theta + u_{\theta t}$$

$$\bar{w}_\theta = \beta_0 + \beta_1 \theta + \dots + \beta_r \theta^r$$

La spécification de SHILLER [1973], un peu plus restrictive, suppose que le terme stochastique $u_{\theta t}$ ne dépend que de θ . Une comparaison de différents estimateurs de ces modèles peut être trouvée dans l'article de ULLAH et RAJ [1978].

6 - CONCLUSION GENERALE

Une première réflexion que l'on peut dégager de cette étude est la précision relativement faible avec laquelle sont estimés simultanément

les coefficients à long terme et la forme des retards échelonnés des modèles. L'incertitude qui affecte les coefficients à long terme est très importante pour le "modélisateur" car la valeur de ces coefficients est souvent -lorsqu'on n'y prête pas attention- la cause d'évolutions divergentes ou aberrantes d'un modèle global. Or ce résultat, loin de reposer sur une "évidence empirique", pourrait être corrigé en fixant à priori, et conformément à la théorie, la valeur de ces coefficients. Dans les exemples traités (demande d'investissement et d'emploi), les hypothèses habituelles sur les fonctions de productions permettent de supposer que l'élasticité du capital et de l'emploi à la production est voisine de l'unité et les estimations des modèles dynamiques, même lorsqu'elles conduisent à des valeurs assez sensiblement différentes, n'ont jamais permis de rejeter cette hypothèse.

L'étude des méthodes d'estimations des modèles à retards échelonnés montre que les conclusions que l'on peut tirer des études empiriques dépendent de façon cruciale des hypothèses à priori concernant la forme des retards et les caractéristiques stochastiques du modèle (notamment pour les modèles estimés sous forme autoregressive). Aussi convient-il d'étudier soigneusement la validité de ces hypothèses en examinant par exemple la sensibilité des résultats à des hypothèses alternatives.

Pour les distributions infinies estimées sous forme autoregressive, la méthode des moindres carrés ordinaires nous paraît, en règle générale, la méthode à retenir, en vérifiant notamment que l'hypothèse d'indépendance des erreurs qui la valide est conforme à la réalité (test de DURBIN). Dans le cas où cette dernière hypothèse est rejetée, la méthode des variables instrumentales (pour les grands échantillons), ou les méthodes applicables à une autocorrélation du premier ordre (COCHRANE-ORCUTT ou HILDRETH-LIU) nous paraissent les plus appropriées aux études empiriques.

En ce qui concerne enfin l'estimation des distributions polynômiales (ALMON) par "balayage" sur le degré (r) et la longueur des délais (s), il faut se souvenir qu'elle a tendance en règle générale, à surestimer ces deux paramètres. En outre, afin de ne pas imposer une contrainte à priori, non conforme à la réalité, il importe de tester la validité de l'hypothèse retenue pour la distribution polynômiale. Le test simple d'analyse de la variance suggéré par DHRYMES devrait à cet égard devenir classique pour ce type de modèle.

REFERENCES BIBLIOGRAPHIQUES

- ANDO A.K., MODIGLIANI F., RASCHE R., TURNOVSKY S.J. (1974)- "On the role of expectations of price and technological change in an investment fonction" - International Economic Review - Vol. 15 - n°2.
- ALLEN R.G.D. (1969)- "Théorie macroéconomique" - Armand Colin
- ALMON S. (1965) - "The Distributed Lag Between Capital Appropriations and Expenditures" - Econometrica - 33 - pp. 178-196.
- ALT F.L. (1942) - "Distributed Lags" Econometrica- 10 -pp. 113-128.
- AMEMIYA T., FULLER W. - "A Comparative Study of Alternative Estimators in a Distributed-Lag Model" - Econometrica - 35 - pp. 509-529. (1967)
- CAGAN P. (1956) - "The Monetary Dynamics of Hyper Inflations" - in Studies in the Quantity Theory of Money, M. Friedman (ed.) - University of Chicago Press, Chicago.
- COCHRANE D., ORCUTT G. (1949) - "Application of Least Squares Regression to Relationships Containing Auto-Correlated Error Terms" - Journal of the American Statistical Association-44-p.32.
- COX D.R. (1961) - "Prediction by Exponentially Weighted Moving Averages and Related Methods" - Journal of the Royal Statistical Society - Série B - 23 - pp. 414-422.
- DURBIN J. (1970) - "Testing for Serial Correlation in Least Squares Regression when Some of the Regressors are Lagged Dependent Variables" - Econometrica - 38 - pp. 410-421.
- DHRYMES P.J. (1971)- "Distributed lags, Problems of Estimation and Formulation" Holden-Day - San Francisco.
- DHRYMES P.J. - KLEIN L.R., STEIGLITZ K. (1970) - "Estimation of Distributed Lags" - International Economic Review - 11 - pp. 235-250.
- EISNER R., STROTZ R. (1963) - "Determinants of Business Investment" - in CMC, Impacts of Monetary Policy, Prentice-Hall, Inc., Englewood Cliffs, New Jersey - pp. 59-337.
- FISHER I. (1957) - "Note on a Short-Cut Method for Calculating Distributed Lags" - Bulletin de l'Institut International de Statistique - 29 - pp. 323-328.
- FOUQUET D., CHARPIN J.M., GUILLAUME M., MUET P.A., VALLET D. (1978) - "Le Modèle DMS" - Collections de l'INSEE - C 64-65.
- FOURGEAUD C., LENCLUD B. - "Econométrie" - PUF - 4ème trimestre . (1978)

- FRIEDMAN M., (1957) - "A Theory of the Consumption Function" - Princeton University Press, Princeton.
- FROST P.A (1975) - "Some properties of the Almon Lag Technique when One Searches for the Degree of Polynomial and Lag" - Journal of the American Statistical Association - Vol. 70 - n°351 - pp. 606-612.
- FULLER W.A. (1962) - "Estimating the Reliability of Quantities Derived from Empirical Production Functions" - Journal of Farm Economics - 44 - pp. 82-99.
- FULLER W.A., MARTIN J.E. (1961) - "The Effects of Autocorrelated Errors on the Statistical Estimation of Distributed Lag Models" - Journal of Farm Economics - 43 - pp. 71-82.
- GODFREY L.G., POSKITT D.S. (1975) - "Testing the Restrictions of the Almon Lag Technique" - Journal of the American Statistical Association - Vol. 70 - n° 349 - pp. 105-108.
- GOULD J.P. (1968) - "Adjustments costs in the theory of investment of the firm" - Review of Economic Studies - Vol. 35 - n° 1.
- GRILICHES Z. (1961) - "A note on the Serial Correlation Bias in Estimates of Distributed Lags" - Econometrica - 29 - pp. 65-73.
- GRILICHES Z. (1967) - "Distributed Lags, A Survey" - Econometrica - 35 - pp. 16-49.
- GRILICHES Z., WALLACE N. (1965) - "The Determinants of Investment Revisited" - International Economic Review - 6 - pp. 311-329.
- HANNAN E.J. (1965) - "The Estimation of Relationships Involving Distributed Lags" - Econometrica - 33 - pp. 206-224.
- HANNAN E.J. (1967) - "The Estimation of a Lagged Regression Relation" - Biometrika - 54 - pp. 409-418.
- HENIN P.Y., PAULRE B., POUCHAIN M. (1976) - "Les comportements d'entreprises comme processus d'ajustement : le modèle RAMAJE" - Association Cournot.
- HILDRETH C., LU J.Y. (1960) - "Demand Relations with Auto-correlated Disturbances" Michigan State University, Agricultural Experimental Station, Technical Bulletin 276.
- HOLLY A. (1978) - "A note on tests of serial correlation based upon instrumental variables estimators" - CEREMADE (University of Paris IX-Dauphine) - Laboratoire d'Econométrie de l'Ecole Polytechnique.

- HOUTHAKKER H.S., TAYLOR L.D. (1970) - "Consumer demand in the United States : Analysis and projections"-Harvard University Press - Cambridge Mass - 2d édition.
- JOHNSTON J. (1972)- " Econometric Methods" - Mac Graw Hill.
- JORGENSON D.W. (1966) - "Rational distributed lags functions" *Econometrica* - Vol. 34 - pp. 135-149.
- KLEIN L.R. (1958) - "The Estimation of Distributed Lags" - *Econometrica* - 26 - pp. 553-56.
- KOYCK L.M. (1954) - "Distributed Lags and Investment Analysis" - North-Holland Publishing Compagny - Amsterdam.
- LIVIATAN N. (1963) - "Consistent Estimation of Distributed Lags" - *International Economic Review* - 4 - pp. 44-52.
- LUCAS R.E. Jr. (1967) - "Optimal Investment Policy and the Flexible Accelerator"- *International Economic Review* - 8 - pp. 78-85.
- MADDALA G.S. (1977) - "Econometrics" - New-York - Mac Graw-Hill.
- MALINVAUD E. (1961-a) - "Estimation et Prévision dans les Modèles Economiques Autoregressifs" - *Review of the International Statistical Institute* - 29 - pp. 1-32.
- MALINVAUD E. (1961-b) - "The Estimation of Distributed Lags : A Comment" - *Econometrica* - 29 - pp. 430-433.
- MALINVAUD E. (1969) - "Méthode Statistiques de l'Econométrie" - Dunod - 2ème édition.
- de MENIL G., YOHN F. (1974) - "La formation de capital fixe des entreprises" *Annales de l'INSEE* n° 26-27 consacré au modèle METRIC.
- MERRIWETHER J.D.(1973) - "Small sample properties of distributed lag estimators with mis-specified lag structure" - *Journal of American Statistical Association* - Vol. 68-n° 343.
- MORRISON J.L., (1970) - "Small Sample Properties of Selected Distributed Lag Estimators : A Monte Carlo Experiment", *International Economic Review*, 11 - pp.13-23.
- MORTENSEN D.T. (1970) - "A theory of wage and employment dynamics" - in PHELPS- "Microeconomic foundations of Employment and inflation theory" - New-York - NORTON.
- MUET P.A. (1976) - "Croissance, profits et investissement" - Communication aux Journées d'Economie et de Finances - 10 décembre 1976.
- MUET P.A. (1978) - "Croissance, profits et investissement : une étude économétrique" - Thèse III cycle Université de Paris I.

- MUET P.A., ZAGAME P. (1976) - "Fonction d'investissement et retards échelonnés" - Annales de l'INSEE n° 21.
- MUNDLAK Y. (1961) - "Aggregation over Time in Distributed Lag Models" - International Economic Review - 2 - pp. 154-163.
- MUNDLAK Y. (1966) - "On the Microeconomic Theory of Distributed Lags" - Review of Economics and Statistics - 48 - pp. 51-60.
- MUTH J.F. (1960) - "Optimal Properties of Exponentially Weighted Forecasts of Time Series with Permanent and Transitory Components" - Journal of the American Statistical Association - 55 - pp. 299-306.
- NADIRI, ROSEN (1969) - "Interrelated factor demand functions" - American Economic Review.
- NADIRI, ROSEN (1974) - "A disequilibrium model of demand for factor of production" - American Economic Review.
- NAGAR A.L., GUPTA Y.P. (1968) - "The Bias of Liviatan's Consistent Estimator in a Distributed Lag Model" - Econometrica - 36 - pp. 337-
- NERLOVE M. (1958) - "Distributed Lags and Demand Analysis for Agricultural and Other Commodities" - U.S Department of Agriculture, Agricultural Marketing Service, Agricultural Handbook - n° 141 - Washington.
- NERLOVE M., WAGE S. (1964) - "On the Optimality of Adaptive Forecasting" - Management Science - 10.
- NERLOVE M. (1968) - "Distributed Lags and Unobserved Components in Economic Time Series" - in W. Fellner (ed.) - Ten Studies in the Tradition of Irving Fisher - John Wiley et Sons, Inc. New York.
- NERLOVE M. (1972) - "Lags in Economic Behavior" - Econometrica - Vol. 40 - n°2.
- POUCHAIN M. (1978) - "Emploi, Capital et taux d'utilisation des facteurs : un essai d'évaluation des hiérarchies d'ajustement" - Université de Paris I - UER - Analyse et Politique Economique.
- SCHMIDT P., WAUD R. (1973) - "The Almon lag technique and the monetary versus fiscal policy debate" - Journal of American Statistical Association - Vol. 68 - n° 341.
- SCHMIDT P. (1974) - "A Modification of the Almon Distributed Lag" - Journal of the American Statistical Association - pp. 679-681.
- SCHMIDT P., SICKLES R. (1975) - "On the Almon lag technique" - I.E.R - Vol. 16 - n° 3.
- SCHRAMM (1970) - "The influence of relative prices, productions, conditions and adjustment costs of investment behavior" - Review of Economics Studies.

- SHILLER R.J. (1973)- "A Distributed Lag Estimator Derived from Smoothness Priors" - *Econometrica* - pp.775-788.
- SOLOW R.M. (1960) - "On a family of lag distributions" - *Econometrica* - Vol. 28 - n° 2.
- THEIL H., STERN R.M. (1960) - "A Simple Unimodal Lag Distribution" - *Metroeconomica* - 12 - pp. 111-119.
- TINSLEY P.A. (1967)- "An application of variable weight distributed lags"- *Journal of the American Statistical Association* - Vol. 62- n° 320.
- TINSLEY P.A. (1979)- "On optimal dynamic adjustment of quasifixed factor" - *F.R.B Spécial studies* - n° 9.
- TREADWAY A.B. (1971)-"The rational multivariate flexible accelerator" - *Econometrica* - Vol. 39 - n° 5.
- ULLAH A., RAJ B. (1978) - "A Polynomial Distributed Lag Model with Stochastic Coefficients" - *Communication au congrès européen de la Société d'Econométrie Genève.*
- VANGREVELINGHE G. (1973) - "Econométrie" - Hermann.