

**Modelos A Coefficients Non-Constants:
Les Principales Formulations**

by

Maximiano Pinheiro
Faculty of Economics
New University of Lisbon("Nova")
July 1992
Working Paper nº 189



TABLE DES MATIERES

1. Introduction
2. Coefficients purement aléatoires
 - 2.1 Le modèle de Hildreth et Houck
 - 2.2 Le modèle de Swamy
3. Coefficients qui sont des fonctions de variables exogènes
 - 3.1 Données exclusivement temporelles ou transversales
 - 3.2 Données mixtes
4. Coefficients qui évoluent selon des processus autorégressifs
 - 4.1 Le modèle de Cooley et Prescott
 - 4.2 Le filtre de Kalman

1. INTRODUCTION

La théorie économique doit être la base indispensable de toute étude d'économie appliquée. Cependant, l'aide qu'elle peut apporter au choix des formes fonctionnelles à utiliser dans ces études appliquées reste au niveau des généralités, de l'encadrement. Dans la plupart des cas, la théorie économique nous donne seulement la liste des variables explicatives possibles et nous dit quels seront les signes des dérivées et/ou élasticités par rapport à ces variables. Excepté des contraintes de convexité/concavité (ou quasi-convexité/quasi-concavité) qui doivent être respectées, l'économètre doit prendre la décision de choisir le type de fonction à estimer se basant sur son expérience, son intuition et aussi sur les résultats qu'il obtient. Souvent, il ne fait pas plus qu'essayer plusieurs spécifications linéaires (dans le sens large de linéarité dans l'espace des

paramètres) et décide après laquelle retenir.

La linéarité dans l'espace des paramètres apparaît justifiée surtout par la simplicité de calcul et par la facilité d'interprétation des résultats. Parfois, on argumente que si la vraie fonction est non-linéaire, la spécification linéaire peut être envisagée comme une approximation locale de cette fonction inconnue. On va même plus loin, raisonnant dans le contexte d'un développement de Taylor, qui permettrait d'interpréter les coefficients de la forme linéaire comme des dérivées de la vraie fonction à un point donné, par exemple le centre de gravité de l'échantillon. En général ceci n'est pas vrai, et bien qu'on puisse admettre l'existence d'un point auquel les estimatives et les dérivées coïncident, localiser ce point demande la connaissance de la vraie spécification non-linéaire de la fonction (voir, par exemple, WHITE (1980)).

L'arbitraire du choix d'une forme linéaire (à coefficients constants) est rendu apparent par une multitude d'arguments qui ont été présentés dans la littérature économétrique depuis le début des années 70. Les plus importants remarquent la non-linéarité due au changement (dans le temps et l'espace) de l'environnement des agents. Si on utilise des séries temporelles pour l'estimation des relations, et si la période d'observation est assez étendue, il est très improbable que les coefficients de réaction des agents n'évoluent pas de façon significative (et graduelle) dans le temps. Si l'environnement change (préférences, produits disponibles, techniques, etc.), il est naturel que les coefficients changent. De façon similaire, si on utilise des données transversales (*cross section*), avec des unités économiques de tailles différentes, il sera aussi très probable que les coefficients de réaction des petites unités ne soient pas les mêmes que ceux

des unités de grande dimension.

L'utilisation de spécifications linéaires à coefficients constants est pourtant discutable et on pourra se douter de la validité des analyses basées sur de telles spécifications. La solution ne se trouve pas dans l'économétrie non-linéaire. Les méthodes non-linéaires supposent connue la forme de la fonction à estimer et le choix d'une certaine spécification non-linéaire est autant arbitraire que le choix d'une autre, dès que les deux vérifient les conditions imposées par la théorie économique pour le problème en étude. En plus, et par rapport aux spécifications linéaires, on perd une bonne partie de la facilité d'interprétation des résultats.

Les tentatives de solution proposées dans la littérature économétrique sont allées dans le sens de choisir des approches plus souples, comme la régression non-paramétrique et les modèles à coefficients non-constants.

Dans cet article nous allons présenter les principales formulations des modèles à coefficients non-constants¹. Ces modèles laissent les coefficients varier selon des règles stochastiques définies *a priori*, tout en gardant une linéarité apparente qui facilite l'interprétation économique.

La bibliographie sur les modèles à coefficients non-constants, toutes versions confondues, est considérable. Même si on trouve des contributions plus anciennes, c'est à partir de la deuxième moitié des années 60 que la

¹PINHEIRO (1990), EUBANK (1988) et HARDLE (1990) sont des exemples de références introductoires sur les méthodes de régression non-paramétrique.

recherche sur le sujet devient plus intense. Dès la fin des années 70, on assiste à une période de consolidation et de sophistication des modèles proposés auparavant: les *surveys* se multiplient et plusieurs manuels d'économétrie incluent des résumés plus ou moins complets sur cette problématique. C'est le cas, par exemple, de MADDALA (1977), CHOW (1983), FOMBY et al (1984), JUDGE et al (1985) et AMEMIYA (1985); HSIAO (1986) comporte aussi un chapitre sur la question sous une perspective de données mixtes (chronologiques/transversales); RAJ et ULLAH (1981) est un exemple d'ouvrage de divulgation entièrement consacré au problème.

Nous n'allons pas faire un *survey* détaillé de tous les articles et livres publiés². Nous n'inclurons pas non plus la vaste littérature sur les *Switching Regression Models* et les *Quantity Rationing Models*, ces derniers aussi appelés de *Disequilibrium Models*, qui admettent la possibilité de changements de régime (de structure) à des observations déterminées (connues ou inconnues). Des résumés et des bibliographies complètes sur ces modèles peuvent être trouvés, par exemple, dans VILARES (1986), MADDALA (1983, 1986) et QUANDT (1982). Nous essayerons plutôt de donner une idée des principaux modèles qui admettent un nombre non-fini de valeurs possibles pour les coefficients. Nous avons divisé ces modèles en trois grandes catégories:

- (i) Modèles à coefficients purement aléatoires (variation aléatoire autour d'une moyenne fixe);
- (ii) Modèles dont les coefficients sont des fonctions linéaires de variables exogènes;

²Une bibliographie annotée assez complète jusqu'à 1980 peut être trouvée dans JOHNSON (1977, 1980).

(iii) Modèles dont les coefficients suivent des processus autorégressifs (stationnaires ou pas).

2. COEFFICIENTS PUREMENT ALEATOIRES

Les modèles à coefficients purement aléatoires ont été les premiers modèles à coefficients non-constants à être proposés dans la littérature. WALD (1947), RUBIN (1950), HURWICZ (1950), KLEIN (1953), THEIL et MENNES (1959), RAO (1965), FISK (1967) et ZELLNER (1966) peuvent être considérés comme les précurseurs de ce type de modèles. Mais c'est avec HILDRETH et HOUCK (1968) et SWAMY (1970, 1971) qu'ils ont atteint une certaine maturité et divulgation.

2.1 Le modèle de HILDRETH et HOUCK

HILDRETH et HOUCK (1968) ont étudié le modèle:

$$y_t = x'_t \beta_t \quad (t=1, \dots, T) \quad (1)$$

avec $x'_t = [1 \ x_{2t} \dots x_{kt}]$ et $\beta_t = \bar{\beta} + v_t$, où v_t est un vecteur aléatoire ($K \times 1$) tel que:

$$E(v_t) = 0, \quad E(v_t v'_t) = \text{diag}(\sigma_1^2, \dots, \sigma_K^2) \equiv D, \quad v_t$$

$$E(v_t v'_s) = 0 \quad (t \neq s) \quad ^3$$

³ Remarquons que si le modèle comportait un bruit blanc u_t ($y_t = x'_t \beta_t + u_t$) il serait impossible d'identifier séparément la variance de β_{1t} et celle de u_t . En effet,

$$y_t = x'_t \beta_t + u_t = (\bar{\beta}_1 + v_{1t}) + x_{2t} (\bar{\beta}_2 + v_{2t}) + \dots + x_{kt} (\bar{\beta}_k + v_{kt}) \quad \text{avec}$$

Ce modèle est une généralisation à plusieurs variables du modèle utilisé par THEIL et MENNES (1959) pour étudier, pour le Royaume Uni, le prix à l'exportation en fonction uniquement du prix à l'importation (sans terme constant et avec coefficient aléatoire).

Substituant $\beta_t = \bar{\beta} + v_t$ dans (1) on obtient le modèle à erreur hétéroscédastique

$$y_t = x_t' \bar{\beta} + w_t \quad (t=1, \dots, T) \quad (2)$$

avec $w_t = x_t' v_t + u_t$, $E(w_t) = 0$, $E(w_t w_s) = 0$ ($t \neq s$) et $E(w_t^2) = x_t' D x_t = \sum_{j=1}^K \sigma_j^2 x_{jt}^2$. Pour estimer σ_j^2 ($j=1, \dots, K$) HILDRETH et HOUCK ont proposé la procédure suivante, en deux étapes:

- 1) La première étape consiste à estimer le modèle $y = X\bar{\beta} + w$ par les moindres carrés ordinaires (m.c.o.), de façon à obtenir le vecteur d'erreurs estimées:

$$\hat{w} = Mw \text{ avec } M = I_T - X(X'X)^{-1}X' = [m_{ts}]$$

$$\hat{w}_t = m_{t1} w_1 + \dots + m_{tT} w_T \quad (t=1, \dots, T)$$

Il est facile de voir que $E(\hat{w}_t) = 0$ et, à cause de la non-corrélation des w_t , que

$$V(\hat{w}_t^2) = m_{t1}^2 V(w_1) + \dots + m_{tT}^2 V(w_T) \quad (3)$$

Si on pose $r = [\hat{w}_t^2]_{(T \times 1)}$, $\dot{M} = [m_{ts}^2]_{(T \times T)}$, $\dot{X} = [x_{tj}^2]_{(T \times K)}$, $Z = \dot{M}\dot{X}$ et $\sigma' = [\sigma_1^2 \dots \sigma_K^2]$,

(3) peut s'écrire $E(r) = Z\sigma$, d'où

$$r = Z\sigma + \xi \quad (4)$$

$$v_{it}^* = v_{it} + u_t$$

avec $E(\xi) = 0$.

2) La deuxième étape consiste à appliquer les m.c.o. à l'équation (4):

$$\hat{\sigma} = (Z'Z)^{-1}Z'r \quad (5)$$

Cet estimateur est non-biaisé et HILDRETH et HOUCK montrent qu'il est aussi convergent.

L'estimateur (5) n'est ni efficace, ni nécessairement non-négatif (comme il conviendrait pour des estimateurs de variances). Plusieurs tentatives de solution de ces problèmes ont été proposées⁴:

(i) Pour satisfaire l'exigence de non-négativité, HILDRETH et HOUCK suggèrent deux possibilités: ou utiliser l'estimateur tronqué $\sigma^* = \max \{0, \hat{\sigma}\}$ (qui ne sera plus centré), ou obtenir l'estimateur de σ comme la solution du programme quadratique

$$\min_{\sigma} (r - Z\sigma)'(r - Z\sigma) \quad \text{sous } \sigma \geq 0$$

(ce qui est plus lourd en calcul et ne résoud pas la question de l'efficacité).

(ii) FROELICH (1973) et SINGH et al (1976) suggèrent d'améliorer l'efficacité, sous l'hypothèse de normalité, par l'utilisation de la méthode du maximum de vraisemblance (m.v.) et ils constatent des problèmes de convergence de la procédure itérative de maximisation.

(iii) Dans le même but d'améliorer l'efficacité, SINGH et al (1976) suggèrent d'estimer σ en 3 étapes: 1) d'abord appliquer (5); 2) après, utiliser

⁴ Les deux problèmes sont liés. En effet, étant donné que les vraies variances sont positives, des estimateurs plus efficaces de σ réduisent la probabilité d'obtenir des estimations négatives.

le résultat de cette première étape pour estimer la matrice de variances-covariances de ξ dans (4); 3) appliquer les m.c.g. sur cette équation. Ils montrent que $E(\xi\xi') = Q$, où $(1/2)Q$ est la matrice dont les éléments sont les carrés des éléments correspondants de MVM et

$$V = V(w) = \text{diag}(\sum_j \sigma_j^2 x_{j1}^2, \dots, \sum_j \sigma_j^2 x_{jT}^2).$$

- (iv) AMEMIYA (1977) propose un estimateur qu'il démontre être aussi efficace asymptotiquement que l'estimateur du m.v.:

$$\hat{\sigma}_A = (\hat{X}' \hat{D}^{-2} \hat{X})^{-1} \hat{X}' \hat{D}^{-2} r \quad (6)$$

où $\hat{D}$ est un quelconque estimateur convergent de D . Bien que l'estimation convergente de D puisse être obtenue appliquant (5), AMEMIYA suggère une méthode alternative inspirée de GOLDFELD et QUANDT (1972, chap.3), plus simple à calculer que (5) parce qu'elle n'exige pas le calcul de $\hat{M}$:

$$\begin{aligned} \hat{w} &= Mw \Leftrightarrow \hat{w}_t = w_t - x_t'(X'X)^{-1}X'w \quad (t=1, \dots, T) \\ \Leftrightarrow r_t &\equiv \hat{w}_t^2 = w_t^2 - 2w_t x_t'(X'X)^{-1}X'w + [x_t'(X'X)^{-1}X'w]^2 \\ \Leftrightarrow r_t &= \sum_j x_{jt}^2 \sigma_j^2 + \eta_t \quad (t=1, \dots, T) \\ \Leftrightarrow r &= \hat{X}\sigma + \eta \end{aligned} \quad (7)$$

avec $\eta_t = \eta_{1t} - 2\eta_{2t} + \eta_{3t}$, $\eta_{1t} = w_t^2 - \sum_j x_{jt}^2 \sigma_j^2 = w_t^2 - E(w_t^2)$, $\eta_{2t} = w_t x_t'(X'X)^{-1}X'w$ et $\eta_{3t} = [x_t'(X'X)^{-1}X'w]^2$. AMEMIYA montre que les m.c.o. appliqués à (7) donnent des estimateurs convergents.

- (v) SRIVASTAVA et al (1981) proposent une méthode d'inspiration bayésienne qui donne aussi des estimateurs asymptotiquement efficaces et qui, en plus, permet de réduire la possibilité d'obtenir des valeurs négatives pour les variances. Pour appliquer cette méthode, on devra commencer par établir *ad hoc* des bornes supérieures pour σ_j^2 :

$$0 \leq \sigma_j^2 \leq q_j \quad (j=1, \dots, K)$$

On traite ces intervalles comme s'ils étaient des intervalles de confiance sous normalité:

$$(1/2)q_j = \sigma_j^2 + v_j \quad (j=1, \dots, K) \quad (8)$$

avec $E(v_j) = 0$ et $E(v_j^2) = (1/2p)q_j$, où p (par exemple: $p = 1, 2$ ou 3) traduit le degré de confiance qu'on accorde aux valeurs q_j proposées.

Combinant (4) et (8) et appliquant les m.c.g. au modèle mixte:

$$\begin{cases} r = Z\sigma + \xi & (E(\xi\xi') = Q) \\ (1/2)q = \sigma + v & (E(vv') = \Psi) \end{cases} \quad (9)$$

où $\Psi = (1/2p)\text{diag}(q_1, \dots, q_K)$, on obtient les estimateurs de σ . On estimera Q utilisant $\hat{V}(w_t) = \text{diag}(\sum_j (1/2)q_j x_j^2)$ ($t=1, \dots, T$) (voir (iii)). Remarquons que dans l'expression de l'estimateur de σ ainsi construit,

$$\hat{\sigma}_M = [Z' \hat{Q}^{-1} Z + \Psi^{-1}]^{-1} [Z' \hat{Q}^{-1} r + (1/2) \Psi^{-1} q] \quad (10)$$

les présences de Ψ^{-1} (matrice diagonale avec tous les éléments positifs) et de q (≥ 0) diminuent considérablement les chances d'obtenir des estimations négatives des variances σ_j^2 . Mais n'oublions pas que cette méthode est applicable uniquement si on est capable d'établir a priori des bornes pour σ_j^2 ($j=1, \dots, K$).

HILDRETH et HOUCK ont traité uniquement le problème d'estimation de la moyenne $\bar{\beta}$ des coefficients β_t . GRIFFITHS (1972) a établi pour la première fois le prévisseur de la trajectoire $\{\beta_t\}_{t=1, \dots, T}$:

$$\hat{\beta}_t = \bar{\beta} + D x_t (x_t' D x_t)^{-1} (y_t - x_t' \bar{\beta}) \quad (11)$$

$\hat{\beta}_t$ est le meilleur prévisseur linéaire non-biaisé de β_t dans le sens que

$E(\hat{\beta}_t - \beta_t) = 0$ et que la matrice

$$\{ E[(\beta_t^* - \beta_t)(\beta_t^* - \beta_t)'] - E[(\hat{\beta}_t - \beta_t)(\hat{\beta}_t - \beta_t)'] \}$$

est définie non-négative pour n'importe quel prévisseur β_t^* linéaire et non-biaisé de β_t .

Une des hypothèses les plus contraignantes du modèle de HILDRETH et HOUCK est sans doute la non-corrélation entre les coefficients des différentes variables de l'équation. SWAMY et MEHTA (1975) relâchent cette hypothèse de diagonalité de la matrice D des variances-covariances des coefficients et proposent, sous l'hypothèse de normalité, trois estimateurs pour δ , le vecteur colonne avec tous les éléments de la diagonale et en dessous de la diagonale de D:

- (i) Une des procédures correspond à la maximisation itérative de la vraisemblance, partant d'une estimation initiale (*initial guess*) δ^* . Les estimateurs ainsi trouvés ont toutes les bonnes propriétés asymptotiques de la méthode du m.v., mais la procédure est assez lourde en calcul.
- (ii) Les deux autres procédures suggérées partent aussi d'une estimation initiale δ^* (SWAMY et MEHTA proposent comme exemple le δ^* correspondant à $D = I_k$). Cette estimation initiale permet d'appliquer les m.c.g. et d'obtenir des erreurs estimées w_t^* ($t=1, \dots, T$). Par des manipulations algébriques, SWAMY et MEHTA réussissent à sortir des estimations de v_t ($\equiv \beta_t - \bar{\beta}$), ce qui leur permet d'utiliser la somme des $\sum_t v_t^*(v_t^*)'$ pour estimer δ . Ils fournissent deux façons alternatives de le faire, avec propriétés comparables: une par résolution d'un système linéaire exact et une autre par régression. Ces estimateurs ne garantissent pas que $\hat{D}$ soit définie non-négative et, en particulier, que les éléments de sa

diagonale soient non-négatifs. Si cette dernière possibilité est vérifiée, SWAMY et METHA suggèrent de tronquer les estimateurs, ou d'utiliser des informations *a priori* éventuellement disponibles sur δ .

Le test de l'hypothèse de non-variabilité des coefficients dans le modèle de HILDRETH et HOUCK, ou dans la généralisation de SWAMY et MEHTA, est équivalent au test d'homoscédasticité de l'erreur du modèle (2). Le test du multiplicateur de Lagrange proposé par BREUSCH et PAGAN (1979) est applicable et a l'avantage d'être très simple. De façon à inclure les deux versions (H.-H. et S.-M.), posons $V(w_t) = h'_t \alpha$, où h_t est un vecteur connu (px1) dont le premier élément est nécessairement 1⁵. L'hypothèse d'homoscédasticité peut être exprimée comme l'hypothèse de nullité de tous les éléments de α à l'exception du premier:

$$H_0 : \alpha_2 = \alpha_3 = \dots = \alpha_p = 0$$

Posons $g_t = \hat{w}_t / (\sum_t \hat{w}_t^2 / T)$, où $\hat{w}_t$ ($t=1, \dots, T$) sont les erreurs de (2) estimées par les m.c.o.. BREUSCH et PAGAN montrent que sous H_0 une partie de la somme $\sum_t (\hat{g}_t - \bar{g})^2$ est asymptotiquement distribuée χ^2_{p-1} , où $\bar{g} = (1/T) \sum_t \hat{g}_t$ et $\hat{g}_t$ ($t=1, \dots, T$) sont les valeurs estimées de g_t par application des m.c.o. à la régression $g_t = h'_t \alpha + \eta_t$.

2.2 Le modèle de SWAMY

Le modèle proposé par SWAMY (1970,1971) est un modèle approprié pour

⁵BREUSCH et PAGAN formulent le problème d'une façon encore plus générale:

$$V(w_t) = \psi(h'_t \alpha).$$

l'utilisation de données mixtes chronologiques/transversales (*panel data*).

On notera y_{it} l'observation de la variable y pour l'"individu" i à l'époque t (et de même pour toutes les autres variables du modèle). La formulation de SWAMY est la suivante:

$$y_{it} = x'_{it} \beta_i + u_{it} \quad (i=1, \dots, N; t=1, \dots, T) \quad (12)$$

avec $\beta_i = \bar{\beta} + v_i$ ($K \times 1$), $E(v_i) = 0$, $E(v_i v_i') = V$ définie positive, $E(v_i v_j') = 0$ ($i \neq j$), $E(u_{it}) = 0$, $E(u_{it}^2) = \sigma_i^2$ et $E(u_{it} u_{js}) = 0$ ($i \neq j$ ou $t \neq s$). Remarquons: 1) l'absence d'un effet temporel dans l'expression des coefficients, c'est-à-dire que les coefficients peuvent varier d'individu à individu mais restent fixes dans le temps; 2) l'hétéroscédasticité transversale de l'erreur u_{it} (sa variance peut être différente d'individu à individu); 3) si $V = \lambda e_1 e_1'$ (λ scalaire et e_1 la première colonne de I_K) et si $\sigma_i^2 = \sigma_1^2$ ($i=1, \dots, N$), le modèle de SWAMY se réduit au modèle à erreurs composées avec seulement effet individuel.

Si on considère toutes les observations temporelles pour un même individu on a:

$$y_i = X_i \beta_i + u_i \quad (13)$$

où $y_i' = [y_{i1} \dots y_{iT}]$, $X_i' = [x_{i1} \dots x_{iT}]$ et $u_i' = [u_{i1} \dots u_{iT}]$. De façon équivalente,

$$y_i = X_i \bar{\beta} + \epsilon_i \quad \text{avec} \quad \epsilon_i = X_i v_i + u_i \quad (14)$$

Empilant pour tous les individus, on obtient:

$$\begin{matrix} y & = & X & \cdot & \bar{\beta} & + & u \\ (NT \times 1) & & (NT \times K) & & (K \times 1) & & (NT \times 1) \end{matrix} \quad (15)$$

avec

$$\begin{matrix} \epsilon & = & \tilde{X} & \cdot & v & + & u \\ (NT \times 1) & & (NT \times NK) & & (NK \times 1) & & (NT \times 1) \end{matrix}$$

où $y' = [y'_1 \dots y'_N]$, $X' = [X'_1 \dots X'_N]$, $u' = [u'_1 \dots u'_N]$, $\epsilon' = [\epsilon'_1 \dots \epsilon'_N]$,
 $v' = [v'_1 \dots v'_N]$ et $\tilde{X} = \text{diag}(X_1, \dots, X_N)$. La matrice des variances-covariances de ϵ est donnée par:

$$\Omega = E(\epsilon\epsilon') = \text{diag}(\Omega_{11}, \dots, \Omega_{NN}) \quad (16)$$

où $\Omega_{11} = E(\epsilon_1\epsilon'_1) = X_1 V X'_1 + \sigma_1^2 I_T$.

Pour V et σ_1^2 ($i=1, \dots, N$) connus, l'estimateur BLUE de $\bar{\beta}$ est celui des m.c.g.:

$$\hat{\bar{\beta}} = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} y = \left(\sum_1 X'_1 \Omega_{11}^{-1} X_1 \right)^{-1} \sum_1 X'_1 \Omega_{11}^{-1} y_1 = \sum_1 W_1 b_1 \quad (17)$$

où $W_1 = \left(\sum_{i=1}^N R_i^{-1} \right)^{-1} R_1^{-1}$, $R_1 = V + \sigma_1^2 (X'_1 X_1)^{-1}$ et $b_1 = (X'_1 X_1)^{-1} X'_1 y_1$.

Remarquons que b_1 est l'estimateur des m.c.o. appliqués aux observations temporelles de l'individu i . L'expression (17) signifie que $\hat{\bar{\beta}}$ est une somme pondérée matricielle des m.c.o. appliqués individu par individu ("m.c.o. individuels"). Cas particulier: si

$$X_1 = \bar{X} \quad \text{et} \quad \sigma_1^2 = \sigma^2 \quad (i=1, \dots, N)$$

alors $R_1 = \bar{R}$ et $\hat{\bar{\beta}} = (1/N) \sum_{i=1}^N b_i$ (moyenne simple des m.c.o. individuels). RAO (1965) avait déjà traité ce cas particulier et était arrivé au même résultat.

SWAMY (1970) propose

$$\hat{\sigma}_1^2 = (y_1 - X_1 b_1)' (y_1 - X_1 b_1) / (T - K) \quad (i=1, \dots, N) \quad (18)$$

comme estimateurs de σ_1^2 ($i=1, \dots, N$). Il montre que ces estimateurs sont centrés et convergents. De même,

$$\hat{V} = S / (N - 1) - \left[\sum_{i=1}^N \hat{\sigma}_i^2 (X'_i X_i)^{-1} \right] / N \quad (19)$$

où $S = \sum_{i=1}^N (b_i - \bar{b})(b_i - \bar{b})'$ et $\bar{b} = (\sum_{i=1}^N b_i)/N$, est un estimateur centré et convergent de V . Cependant, la matrice $\hat{V}$ peut ne pas être définie non-négative pour des échantillons de taille finie. SWAMY (1971) suggère d'utiliser simplement

$$\hat{V} = S/(N - 1) \quad (20)$$

qui est aussi un estimateur convergent (mais biaisé) de V .

Pour tester

$$H_0 : \beta_1 = \dots = \beta_N = \bar{\beta} \quad (\Leftrightarrow V = 0) \quad (21)$$

SWAMY (1970) propose d'utiliser la statistique

$$h = \sum_{i=1}^N [(b_i - \tilde{\beta})' X_i' X_i (b_i - \tilde{\beta}) / \sigma_i^2] \quad (22)$$

où $\tilde{\beta}$ est l'estimateur contraint de $\bar{\beta}$:

$$\tilde{\beta} = [\sum_{i=1}^N (\hat{\sigma}_i^2)^{-1} X_i' X_i]^{-1} \sum_{i=1}^N (\hat{\sigma}_i^2)^{-1} X_i' y_i \quad (23)$$

Sous H_0 , $h \xrightarrow{d} \chi_{K(M-1)}^2$. HSIAO (1986, p.135) généralise le test de BREUSCH et PAGAN (1979) et propose un test alternatif au précédent test de SWAMY.

Soit le modèle des moyennes individuelles divisées par σ_i :

$$(1/\sigma_i) \bar{y}_i = (1/\sigma_i) \bar{x}_i' \beta_i + (1/\sigma_i) \bar{u}_i \quad (i=1, \dots, N) \quad (24)$$

où $\bar{y}_i = (\sum_t y_{it})/T$, etc.. Le modèle (24) est équivalent à

$$(1/\sigma_i) \bar{y}_i = (1/\sigma_i) \bar{x}_i' \bar{\beta} + \omega_i \quad (i=1, \dots, N) \quad (25)$$

avec $\omega_i = (1/\sigma_i) \bar{u}_i + (1/\sigma_i) \bar{x}_i' v_i$. Etant donné que

$$V(\omega_i) = 1/T + (1/\sigma_i^2) \bar{x}_i' V \bar{x}_i \quad (i=1, \dots, N) \quad (26)$$

⁶ Continuant à admettre $\sigma_i^2 \neq \sigma_j^2$ ($i \neq j$).

tester $V = 0$ se réduit à tester l'homoscédasticité des ω_i (c'est-à-dire, à tester si $V(\omega_i) = 1/T$ pour tout i). Soit la régression

$$(T\hat{\omega}_i^2 - 1) = (1/\hat{\sigma}_i^2) \left[\sum_{j=1}^K \bar{x}_{ji} \bar{x}_{ji} \theta_{ji} \right] + \xi_i \quad (i=1, \dots, N) \quad (27)$$

où les $\hat{\sigma}_i^2$ sont donnés par (18) et les $\hat{\omega}_i$ ($i=1, \dots, N$) sont les erreurs de (25) estimées par les m.c.o. (avec σ_i^2 substitué par $\hat{\sigma}_i^2$). Sous H_0 , une demié de la somme des carrés des estimations de $(T\hat{\omega}_i^2 - 1)$, obtenues par application des m.c.o. à (27), suit asymptotiquement une loi $\chi_{K(K+1)/2}^2$, quand T et N tendent simultanément vers infini.

Pour l'estimation des coefficients individuels β_i ,

$$\hat{\beta}_i = \hat{\beta} + VX'_i(X_i VX'_i + \sigma_i^2 I_T)^{-1}(y_i - X_i \hat{\beta}) \quad (28)$$

est BLUP ("Best Linear Unbiased Predictor") de β_i dans le sens que $E(\hat{\beta}_i - \beta_i) = 0$ et que la matrice $[V(\hat{\beta}_i) - V(\beta_i)]$ est définie non-négative pour tout prévisseur linéaire non-biaisé β_i^* ⁷.

SWAMY (1974) a généralisé son modèle pour les cas où il existerait des variables endogènes décalées et des erreurs autocorrélées. Une autre généralisation qui a été introduite dans la littérature concerne la variabilité des coefficients dans le temps. Dans le modèle original de SWAMY (1970,1971), et comme on l'a remarqué, les coefficients sont constants dans le temps pour un individu donné. HSIAO (1974,1975) a proposé la formulation

⁷ Il semble que ce prévisseur ait été établi pour la première fois par LEE et GRIFFITS (1979) "The Prior Likelihood and Best Linear Unbiased Prediction in Stochastic Coefficient Linear Models", University of New England Working Papers, No 1.

$$y_{it} = x'_{it}(\bar{\beta} + v_i + \lambda_t) + u_{it} \quad (29)$$

avec $E(\lambda_t) = 0$ et $E(\lambda_t \lambda'_t) = \Lambda$ (diagonale), qui permet aux coefficients de varier (ensemble) avec le temps. L'effet λ_t rend, dans ce sens, la version de HSIAO plus générale que celle de SWAMY. Mais, par contre, la version de HSIAO est beaucoup plus exigeante en calcul, même imposant l'homocédasticité de l'erreur u_{it} ($E(uu') = \sigma^2 I_{NT}$) et la diagonalité de V . SWAMY et MEHTA (1975, 1977) ont développé une version plus générale que celle de HSIAO, où Λ_i peut varier d'individu à individu et ni V , ni Λ_i ($i=1, \dots, N$) ne sont plus supposées diagonales. Cependant, elle exige, tout comme la version de HSIAO, un grand volume de calcul.

3. COEFFICIENTS QUI SONT DES FONCTIONS DE VARIABLES EXOGENES

Une généralisation directe des modèles à coefficients purement aléatoires consiste à substituer les moyennes fixes $\bar{\beta}$, autour desquelles les coefficients peuvent varier, par des fonctions linéaires (dans l'espace des paramètres) de variables exogènes (les mêmes que celles du modèle de base ou autres). SINGH et al (1976) désignent ces modèles par "Variable Mean Response Models", par opposition aux modèles à coefficients purement aléatoires, qu'ils appellent "Constant Mean Response Models".

3.1 Données exclusivement temporelles ou transversales

Quand les données ne sont pas mixtes, l'estimation de modèles dont les coefficients sont des fonctions linéaires de variables exogènes se réduit:

- (i) soit à l'estimation du modèle à coefficients constants si les coefficients ne comportent pas de terme aléatoire;

(ii) soit, dans le cas affirmatif, à l'estimation du modèle de l'alinéa 2.1.

Posons:

$$y_t = x'_t \beta_t + u_t \quad (30a)$$

$$\beta_t = Z_t \gamma + v_t \quad (t=1, \dots, T) \quad (30b)$$

$(K \times 1) \quad (K \times m) \quad (m \times 1) \quad (n \times 1)$

où γ est un vecteur de paramètres (inconnus)⁸. Substituant (30b) dans (30a) on obtient

$$y_t = h'_t \gamma + w_t \quad (31)$$

où $h_t = x'_t Z_t$ et $w_t = x'_t v_t + u_t$. Si $v_t \equiv 0$, alors $w_t = u_t$ et (31) se réduit à une formulation à coefficients constants. Si v_t est un vecteur aléatoire d'espérance nulle et matrice de variances-covariances D , (32) est presque la formulation discutée à l'alinéa 2.1. Une seule particularité: le vecteur des variables explicatives dans (31) est $h_t = Z'_t x_t$ et non simplement x_t (mais attention: Z_t est absente de l'expression de l'erreur w_t).

Considérons maintenant le problème du choix des variables explicatives des coefficients à inclure dans Z_t . Dans certains cas, la théorie économique, ou notre intuition, nous dirons lesquelles considérer. Un exemple simple: si on veut estimer une fonction de consommation avec des données transversales, on pourra exprimer le coefficient du revenu comme fonction du revenu lui-même (avec une constante), ce qui permettra de traduire l'éventuelle diminution de la propension marginale à consommer quand le revenu augmente. Pour les séries chronologiques, et quand il n'y a pas un choix naturel de variables explicatives, SINGH et al (1976) suggèrent

⁸On pourra substituer l'indice t par i ($i=1, \dots, N$) sans changer quoi que soit.

d'utiliser des fonctions de "t":

"[...] the typical regression coefficient $\beta_j(t)$ is subject to two influences that cause it to deviate from its average $\bar{\beta}_j$. The first of these, following HILDRETH-HOUCK, is a random disturbance that possesses certain distributional properties. The second is due to the influence of factors that may vary systematically with time. Thus in time series analysis it is likely that response coefficients fluctuate around some trend rather than about the average value $\bar{\beta}_j$. For instance, consumer's taste may vary over time and this may cause systematic shifts in the marginal propensity to consume. More generally, some parameters may rise systematically in the upswing of the business cycle and fall during the down-swing. Should the movement of the parameters in the two phases be symmetric, the path of the parameter over the complete cycle can be approximated by a parabolic trend, or some other suitable trend that can be specified."

(in SINGH *et al* (1976, p.341))

Pour faciliter le choix des variables à inclure dans Z_t (surtout s'il y a un nombre assez large de candidates), BELSLEY (1973) a essayé de séparer en deux phases distinctes et successives l'estimation de (30a) et (30b). L'idée est la suivante: si on pouvait trouver une façon d'estimer β_t indépendamment du choix de Z_t , on pourrait régresser ensuite (et séparément) $\hat{\beta}_t$ en fonction de plusieurs choix possibles de Z_t et retenir le meilleur. Voici la procédure proposée par BELSLEY:

- 1) Fixons τ entre 1 et $T-K$. Utilisant seulement les $K+\tau$ premières observations, on calcule (par les m.c.o.):

$$b_{K+\tau} = (X_{K+\tau}^{*'} X_{K+\tau}^*)^{-1} X_{K+\tau}^{*'} y_{K+\tau}^* \quad (32)$$

où $X_{K+\tau}^* = [x_1 \dots x_{K+\tau}]$ et $y_{K+\tau}^* = [y_1 \dots y_{K+\tau}]$. Ensuite, on laisse tomber la première observation (x_1, y_1) en ajoutant, pour la substituer, la $(K+\tau+1)$ -ème, et on estime:

$$b_{K+\tau+1} = (X_{K+\tau+1}^* X_{K+\tau+1}^{*'})^{-1} X_{K+\tau+1}^* y_{K+\tau+1}^* \quad (33)$$

On continue la procédure jusqu'à l'épuisement des observations (on calculera au total $[T-(K+\tau)+1]$ vecteurs b_i).

2) Dans une deuxième étape, et ayant comme variable dépendante b_i ($i=K+\tau, \dots, T$), on essayera plusieurs spécifications de Z_i :

$$b = Z_* \gamma + v \quad (\hat{\gamma} = (Z_*' Z_*)^{-1} Z_*' b) \quad (34)$$

où $Z_*' = [Z_{K+\tau}' \dots Z_T']$ et $b' = [b_{K+\tau}' \dots b_T']$. On retiendra celle qui le mieux explique b .

Quelques commentaires s'imposent: 1) BELSLEY utilise uniquement l'information passée pour estimer chaque b_i ; 2) en plus, il donne le même poids à toutes les $K+\tau$ observations utilisées (quand il serait plus logique de donner plus d'importance aux observations plus "proches"); 3) finalement, le choix de τ n'est pas résolu. Si on choisit τ très petit, les b_i seront imprécis (à cause du manque de degrés de liberté), et si on le choisit grand, alors la valeur $[T-(K+\tau)+1]$ sera petite et les b_i seront identiques ou presque (et l'étape 2 n'aura pas de sens).

3.2 Données mixtes

Soit la spécification:

$$y_{it} = x_{it}' \beta_i + u_{it} \quad (i=1, \dots, N; t=1, \dots, T) \quad (35a)$$

$$\beta_i = Z_i \gamma + v_i \quad (35b)$$

$(K \times 1) \quad (K \times m) \quad (m \times 1) \quad (K \times 1)$

Substituant (35b) dans (35a) on obtient:

$$y_{it} = x'_{it} Z_1 \gamma + \varepsilon_{it} \quad (i=1, \dots, N; t=1, \dots, T) \quad (36)$$

où $\varepsilon_{it} = x'_{it} v_1 + u_{it}$. Acceptant les mêmes hypothèses sur les termes aléatoires u_{it} et v_1 que dans le modèle de SWAMY (1970, 1971)

$$E(u_{it}) = 0, \quad E(u_{it} u_{is}) = \begin{cases} \sigma^2 & \text{si } i=1 \text{ et } t=s \\ 0 & \text{sinon} \end{cases}, \quad E(v_1 v'_1) = \begin{cases} V & \text{si } i=1 \\ 0 & \text{sinon} \end{cases}$$

l'estimation de (36) peut se faire, à quelques détails près, par la procédure indiquée dans l'alinéa 2.2:

(i) Dans l'expression (17) il faudra substituer $W_1 = (\sum_1 R_1^{-1})^{-1} R_1^{-1}$ par

$$W_1 = (\sum_1 Z'_1 R_1^{-1} Z_1)^{-1} Z'_1 R_1^{-1};$$

(ii) Les m.c.o. devront être appliqués aux équations (36) et non aux équations (35a) ($b_1 = (Z'_1 X'_1 X_1 Z_1)^{-1} Z'_1 X'_1 y_1$).

AMEMIYA (1978) fournit des estimateurs pour une formulation et un cadre d'hypothèses un peu plus général que le précédent, quand il y a des paramètres qui restent constants et quand le terme résiduel u_{it} est corrélé contemporanément pour des individus différents:

$$y_{it} = X_{11} \beta_1 + X_{21} \beta_{21} + u_{it} \quad (i=1, \dots, N) \quad (37a)$$

(Tx1) (TxK₁) (K₁x1) (TxK₂) (K₂x1) (Tx1)

$$\beta_{21} = Z_1 \gamma + v_1 \quad (37b)$$

(K₂x1) (K₂xm) (mx1) (K₂x1)

On suppose les mêmes hypothèses sur v_1 que pour le modèle (35), mais avec:

$$E(u_{it} u_{is}) = \begin{cases} \sigma_{11} & \text{si } t = s \\ 0 & \text{sinon} \end{cases}$$

$$\Leftrightarrow E(u_1 u'_1) = \sigma_{11} I_T \quad (i, l=1, \dots, N)$$

Comme application possible de cette formulation, AMEMIYA suggère le problème traité par WACHTER (1970). Cet auteur avait utilisé *panel data* pour une étude de salaires où les individus étaient des industries américaines (pour

lesquelles il disposait de données annuelles allant de 1947 à 1967). WACHTER a appliqué les m.c.o. individuels à chaque industrie, prenant comme variable expliquée le logarithme (log) du salaire relatif de l'industrie i , et comme variables explicatives (en plus d'un terme constant):

le log de la valeur ajoutée relative de l'industrie i

le log du taux d'inflation

le log du taux de chômage pour l'industrie i

un "trend"

une variable muette pour la période de la guerre de Corée.

Etant donné que le coefficient du log du taux de chômage dépend du degré de concentration de l'industrie et du pourcentage de travailleurs syndicalisés dans chaque industrie, WACHTER a considéré les coefficients estimés du taux de chômage en fonction de ces 2 variables et d'un terme constant, et a appliqué à nouveau les m.c.o.. La formulation (37) de AMEMIYA peut être envisagée comme une solution pour la perte de degrés de liberté provoquée par l'utilisation des m.c.o. individuels, mais surtout elle permet d'explicitier que les industries ne sont pas indépendantes les unes des autres (c'est pourquoi AMEMIYA considère des corrélations contemporaines pour les erreurs u_{it}).

Ecrivons matriciellement le modèle (37) et les hypothèses respectives sur les termes aléatoires:

$$\begin{cases} y = X_1 \beta_1 + X_2 \beta_2 + u & (38a) \\ \beta_2 = Z\gamma + v & (38b) \end{cases}$$

avec $y' = [y'_1 \dots y'_N]$, $X'_1 = [x'_{11} \dots x'_{1N}]$, $X_2 = \text{diag}(X_{21}, \dots, X_{2N})$, $\beta'_2 = [\beta'_{21} \dots \beta'_{2N}]$, $Z' = [Z'_1 \dots Z'_N]$, $u' = [u'_1 \dots u'_N]$ et $v' = [v'_1 \dots v'_N]$. En plus:

$$E(uu') = \Sigma \otimes I_T \quad \text{où} \quad \Sigma_{(N \times N)} = [\sigma_{ij}]$$

$$E(vv') = I_N \otimes V$$

Après substitution de (38b) dans (38a) on obtient:

$$y = X_1 \beta_1 + X_2 Z \gamma + \varepsilon \quad (39)$$

où $\varepsilon = X_2 v + u$ et

$$\Omega \equiv E(\varepsilon \varepsilon') = X_2 (I_N \otimes V) X_2' + (\Sigma \otimes I_T) \quad (40)$$

Pour estimer Σ et V AMEMIYA suggère les estimateurs convergents suivants:

$$\hat{\sigma}_{11} = (y_1 - X_{11} \tilde{\beta}_1 - X_{21} \tilde{\beta}_2)' (y_1 - X_{11} \tilde{\beta}_1 - X_{21} \tilde{\beta}_2) / T \quad (41)$$

$$\hat{V} = (1/N) \sum_{i=1}^N (\tilde{\beta}_{2i} - Z_1 \tilde{\gamma}) (\tilde{\beta}_{2i} - Z_1 \tilde{\gamma})' \quad (42)$$

où

$$\begin{bmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \end{bmatrix} = \begin{bmatrix} X_1' X_1 & X_1' X_2 \\ X_2' X_1 & X_2' X_2 \end{bmatrix}^{-1} \begin{bmatrix} X_1' y \\ X_2' y \end{bmatrix} \quad 9$$

$$\tilde{\gamma} = (Z' Z)^{-1} Z' \tilde{\beta}_2 \quad 10$$

Ayant estimé Σ et V , l'estimation de γ dans (39) peut se faire par application des m.c.g. (en forme partagée):

$$\begin{aligned} \hat{\gamma} = & [Z' X_2' \hat{\Omega}^{-1} X_2 Z - Z' X_2' \hat{\Omega}^{-1} X_1 (X_1' \hat{\Omega}^{-1} X_1)^{-1} X_1' \hat{\Omega}^{-1} X_2 Z]^{-1} \\ & [Z' X_2' \hat{\Omega}^{-1} - Z' X_2' \hat{\Omega}^{-1} X_1 (X_1' \hat{\Omega}^{-1} X_1)^{-1} X_1' \hat{\Omega}^{-1}] y \end{aligned} \quad (43)$$

L'expression précédente n'est pas très convenable parce qu'elle exige

⁹M.c.o. individuels sur (38a).

¹⁰M.c.o. sur (38b) après avoir substitué β_2 par $\tilde{\beta}_2$.

l'inversion de $\hat{\Omega}$ (d'ordre NT). AMEMIYA propose une procédure en 2 étapes, qu'il démontre être équivalente à (43) et qui n'exige pas l'inversion de Ω , mais celle de Σ (d'ordre N):

1) On calcule d'abord

$$\beta_2^* = (X_2' \Sigma^* X_2)^{-1} X_2' \Sigma^* y \quad (44)$$

où $\Sigma^* = (\hat{\Sigma}^{-1} \otimes I_T) - (\hat{\Sigma}^{-1} \otimes I_T) X_1 [X_1' (\hat{\Sigma}^{-1} \otimes I_T) X_1]^{-1} X_1' (\hat{\Sigma}^{-1} \otimes I_T)^{11}$. Remarquons que si on substitue (2.38a) dans (2.44) on obtient (parce que $\Sigma^* X_1 = 0$ et $\beta_2 = Z\gamma + v$):

$$\beta_2^* = Z\gamma + \xi \quad (45)$$

avec $\xi = (X_2' \Sigma^* X_2)^{-1} X_2' \Sigma^* u + v$. La matrice des variances-covariances de ξ est la suivante:

$$E(\xi\xi') = (X_2' \Sigma^* X_2)^{-1} X_2' \Sigma^* (\Sigma \otimes I_T) \Sigma^* X_2 (X_2' \Sigma^* X_2)^{-1} + (I_N \otimes V)$$

Estimant Σ par $\hat{\Sigma}$ et V par $\hat{V}$, et puisque $\Sigma^* (\hat{\Sigma} \otimes I_T) \Sigma^* = \Sigma^*$, la matrice précédente se réduit à $[(X_2' \Sigma^* X_2)^{-1} + (I_N \otimes V)]$, dont l'inverse est:

$$X_2' \Sigma^* X_2 [X_2' \Sigma^* \hat{\Omega} \Sigma^* X_2]^{-1} X_2' \Sigma^* X_2$$

2) La deuxième étape consiste simplement à appliquer les m.c.g. à (45):

$$\gamma^* = \{Z' X_2' \Sigma^* X_2 [X_2' \Sigma^* \hat{\Omega} \Sigma^* X_2]^{-1} X_2' \Sigma^* X_2 Z\}^{-1} Z' X_2' \Sigma^* X_2 [X_2' \Sigma^* \hat{\Omega} \Sigma^* X_2]^{-1} X_2' \Sigma^* X_2 \beta_2^* \quad (46)$$

4. COEFFICIENTS QUI EVOLUENT SELON DES PROCESSUS AUTOREGRESSIFS

A côté des modèles à coefficients purement aléatoires et des modèles

¹¹ Moindres carrés individuels, cette fois généralisés et en forme partagée, sur (38a).

dont les coefficients sont des fonctions de variables exogènes, la troisième grande classe de modèles à coefficients non-constants est constituée par des modèles dont les coefficients sont supposés évoluer (exclusivement) dans le temps selon des processus stochastiques (stationnaires ou pas) choisis *a priori*.

4.1 Le modèle de COOLEY et PRESCOTT

Nous allons d'abord considérer le modèle proposé par COOLEY et PRESCOTT (1973a, 1976). C'est un modèle pour données chronologiques qui est caractérisé par la séparation des coefficients en deux composantes: permanente et transitoire.

"If knowledge of exogenous events is imperfect, theory suggests that it is appropriate to view behavioral relationships as changing slowly over time. Many of the kinds of misspecification [...] imply changes in parameters that are permanent in time. The structure we are assuming [...] implies that the permanent components of the parameter will drift systematically over time away from their initial value with no inherent tendency to return to a mean value."

(in COOLEY et PRESCOTT (1976, p.169))

COOLEY et PRESCOTT admettent que la composante transitoire est purement aléatoire et que la composante permanente évolue selon une promenade aléatoire¹²:

¹²COOLEY et PRESCOTT ont commencé (1973b, 1973c) par développer un cas particulier, qu'ils ont désigné par "Adaptive Regression Model", avec tous

$$\begin{cases} y_t = x_t' \beta_t & (47a) \\ \beta_t = \beta_t^p + w_t & (47b) \\ \beta_t^p = \beta_{t-1}^p + v_t & (47c) \end{cases} \quad (t=1, \dots, T)$$

avec x_t ($K \times 1$), β_t ($K \times 1$), $w_t \sim N(0, (1-\gamma)\sigma^2 \Sigma_w)$, $v_t \sim N(0, \gamma\sigma^2 \Sigma_v)$, $E(w_t w_s') = 0$ ($t \neq s$), $E(v_t v_s') = 0$ ($t \neq s$) et $E(w_t v_s') = 0 \quad \forall_{t,s}$. Les matrices Σ_w et Σ_v ($K \times K$) sont supposées connues¹³ et σ^2 et γ sont deux paramètres à estimer. Si $\gamma = 0$ on retombe sur le modèle à coefficients purement aléatoires. Le paramètre γ traduit l'importance relative des variations permanente et transitoire. En effet, par simple manipulation de (47b) et (47c), $\beta_t = \beta_{t-1} + v_t + (w_t - w_{t-1})$, d'où:

$$V(\beta_t | \beta_{t-1}, y_{t-1}) = \sigma^2 [\gamma \Sigma_v + (1-\gamma) \Sigma_w] \quad (48)$$

L'application de la méthode du maximum de vraisemblance pour estimer simultanément γ , σ^2 et β_t^p ($t=1, \dots, T$) n'est pas envisageable parce que le processus qui engendre β_t^p n'est pas stationnaire. Cependant, si on veut estimer uniquement γ , σ^2 et β_t^p pour une valeur de t fixée, la fonction de vraisemblance conditionnelle à une réalisation particulière de β_t^p est bien définie (et correspond à traiter le β_t^p choisi comme un vecteur fixe de paramètres). COOLEY et PRESCOTT suggèrent de prendre β_{T+1}^p , ce qui est très convenable pour faire de la prévision. Par substitutions successives dans

les paramètres fixes à l'exception de celui correspondant au terme constant.

¹³ Quand le premier élément du vecteur x_t est le terme constant, COOLEY et PRESCOTT suggèrent de normaliser ces deux matrices posant leurs éléments (1,1) égaux à 1. Avec cette normalisation, la composante transitoire de la variation du coefficient du terme constant correspond au terme résiduel du modèle de régression conventionnel à coefficients constants.

(47c),

$$\begin{aligned}
 \beta_{T+1}^p &= \beta_t^p + \sum_{j=t+1}^{T+1} v_j \\
 \Rightarrow \beta_t^p &= \beta_{T+1}^p - \sum_{j=t+1}^{T+1} v_j + w_t \\
 \Rightarrow y_t &= x_t' \beta_{T+1}^p + (x_t' w_t - x_t' \sum_{j=t+1}^{T+1} v_j) \quad (t=1, \dots, T) \\
 \Rightarrow y &= X \beta_{T+1}^p + \varepsilon
 \end{aligned} \tag{49}$$

avec $\varepsilon \sim N(0, \sigma^2 \Omega(\gamma))$, $\Omega(\gamma) = (1-\gamma)R + \gamma Q$, $R = \text{diag}(r_{tt})$, $r_{tt} = x_t' \Sigma_w x_t$, $Q = [q_{ts}]$ et $q_{ts} = \{\min(T-t+1, T-s+1)\} x_t' \Sigma_v x_s$. Les estimateurs du m.v. de β_{T+1}^p et de σ^2 , en fonction de γ , sont donnés par:

$$\hat{\beta}_{T+1}^p(\gamma) = [X' \Omega^{-1}(\gamma) X]^{-1} X' \Omega^{-1}(\gamma) y \tag{50}$$

$$\hat{\sigma}^2(\gamma) = (y - X \hat{\beta}_{T+1}^p)' \Omega^{-1}(\gamma) (y - X \hat{\beta}_{T+1}^p) / T \tag{51}$$

Pour estimer γ , la fonction log-vraisemblance concentrée,

$$L_c(y, \gamma) = \text{cte} - (T/2) \ln \hat{\sigma}^2(\gamma) - (T/2) \ln |\Omega(\gamma)| \tag{52}$$

devra être maximisée, par exemple par balayage dans le domaine admissible de γ ($0 \leq \gamma \leq 1$)¹⁴.

COOLEY et PRESCOTT (1976) montrent que les estimateurs ainsi obtenus sont convergents et asymptotiquement efficaces, et que $\hat{\beta}_{T+1}^p$ permet de prévoir y_{T+1} de façon optimale. Des hypothèses sur γ peuvent être testées utilisant la variance asymptotique de $\hat{\gamma}$:

$$V[\sqrt{T}(\hat{\gamma} - \gamma)] = 2 \left\{ \frac{1}{T} \sum_1 \frac{(d_1 - 1)^2}{d_1(\gamma)^2} - \left[\frac{1}{T} \sum_1 \frac{d_1 - 1}{d_1(\gamma)} \right]^2 \right\}^{-1} \tag{53}$$

¹⁴ COOLEY et PRESCOTT proposent une transformation de (50) qui évite d'inverser $\Omega(\gamma)$ à chaque évaluation de (52).

où $d_i(\gamma) = (1-\gamma) + \gamma d_i$ et les d_i ($i=1, \dots, T$) sont les valeurs propres de $Q^* = [q_{ij}^*]$, avec $q_{ij}^* = q_{ij} / (\sqrt{r_{ii} r_{jj}})$.

4.2 Le filtre de Kalman

La formulation de COOLEY et PRESCOTT peut être envisagée comme un cas particulier d'une formulation plus générale:

$$y_t = x_t' \beta_t + u_t \quad (54a)$$

$$\beta_t = H \beta_{t-1} + v_t \quad (t=1, \dots, T) \quad (54b)$$

$(K \times K)$

Si on pose $\beta_t = \beta_t^p$, $H = I_K$ et $u_t = x_t' w_t$, on retrouve la formulation (47).

Pour $v_t \equiv 0$ et $H = I_K$, (54) se réduit au modèle à coefficients constants.

Considérons les hypothèses suivantes sur les termes aléatoires u_t et v_t :

$$u_t \sim \text{n.i.d.}(0, \sigma^2), \quad v_t \sim \text{n.i.d.}(0, V), \quad E(u_t v_t) = 0 \quad (55)$$

où n.i.d. signifie "normal and independently distributed". Le filtre de Kalman (KALMAN (1960) et KALMAN et BUCY (1961)) constitue la base pour calculer les estimateurs du maximum de vraisemblance de (54) sous les hypothèses (55).

Nous utiliserons la notation:

$\mathcal{I}_t$, ensemble de l'information disponible à l'époque t

$$\beta_{t|s} \equiv E(\beta_t | \mathcal{I}_s)$$

$$\Sigma_{t|s} \equiv \text{cov}(\beta_t | \mathcal{I}_s) = E[(\beta_t - \beta_{t|s})(\beta_t - \beta_{t|s})']$$

L'évaluation de $\beta_{t|t}$ est connue sous le nom de *filtrage*; l'évaluation de

$\beta_{t|s}$ pour $t > s$ est appelée *prévision*; finalement, l'évaluation de $\beta_{t|s}$ pour

$t < s$ est désignée par lissage (*smoothing*).

De (54b) on obtient directement que

$$\beta_{t|t-1} = H\beta_{t-1|t-1} \quad (56)$$

Par ailleurs, d'après (54b) et par application des propriétés des espérances conditionnelles:

$$\beta_{t|t} = \beta_{t|t-1} + E(\beta_t - \beta_{t|t-1} | y_t - x_t' \beta_{t|t-1}) \quad (57)$$

A cause de la normalité des erreurs u_t et v_t , le terme $E(\beta_t - \beta_{t|t-1} | y_t - x_t' \beta_{t|t-1})$ est linéaire en $(y_t - x_t' \beta_{t|t-1})$:

$$\beta_{t|t} = \beta_{t|t-1} + \underset{(k_t)}{k_t'} (y_t - x_t' \beta_{t|t-1}) \quad (58)$$

Le vecteur k_t a été déduit par KALMAN (1960):

$$k_t = \Sigma_{t|t-1} x_t (x_t' \Sigma_{t|t-1} x_t + \sigma^2)^{-1} \quad (59)$$

$$\Sigma_{t|t-1} = H \Sigma_{t-1|t-1} H' + V \quad (60)$$

$$\Sigma_{t|t} = \Sigma_{t|t-1} - k_t (x_t' \Sigma_{t|t-1} x_t + \sigma^2) k_t' \quad (61)$$

Pour σ^2 , H , V , $\Sigma_{0|0}$ donnés, si on utilise récursivement (dans l'ordre) (60)-(56)-(59)-(58)-(61), on peut évaluer $\Sigma_{t|t-1}$, $\beta_{t|t-1}$, $\Sigma_{t|t}$ et $\beta_{t|t}$ pour t allant de 1 à T . Après le filtrage, les vecteurs $\beta_{t|t+n}$ ($n \geq 1$) peuvent être calculés par une autre procédure récursive analogue:

$$\beta_{t|t+n} = \beta_{t|t+n-1} + S_t (\beta_{t+1|t+n} - \beta_{t+1|t+n-1}) \quad (62)$$

$$S_t = \Sigma_{t|t} H' \Sigma_{t+1|t}^{-1} \quad (63)$$

La matrice des variances-covariances $\Sigma_{t|t+n}$ est aussi évaluée récursivement utilisant l'équation:

$$\Sigma_{t|t+n} = \Sigma_{t|t+n} + S_t S_{t-1} \dots S_{t+n-1} (\Sigma_{t+n|t+n} - \Sigma_{t+n|t+n-1}) S_{t+n-1}' \dots S_t' \quad (64)$$

Les démonstrations des résultats précédents peuvent être trouvées, par

exemple, dans SAGE et MELSA (1971) ou CHOW (1981,1983,1984).

Pour H non-singulière, SANT (1977) a montré que les résultats précédents peuvent être déduits dans le cadre de la formulation de Aitken (c'est-à-dire, par l'application de la méthode des m.c.g.)¹⁵. Par substitutions successives dans (54b),

$$\beta_t = H^{t-s} \beta_s + H^{t-(s+1)} v_{s-1} + \dots + H v_{t-1} + v_t \quad (66)$$

Si H^{-1} existe, on peut exprimer $\beta_1, \beta_2, \dots, \beta_s, \dots, \beta_{t-1}$ en fonction de β_t , substituer le résultat dans (54a) et obtenir (pour les t premières observations):

$$\begin{aligned} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{t-1} \\ y_t \end{bmatrix} &= \begin{bmatrix} x_1' H^{-t+1} \\ x_2' H^{-t+2} \\ \vdots \\ x_{t-1}' H^{-1} \\ x_t' \end{bmatrix} \beta_t + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{t-1} \\ u_t \end{bmatrix} - \begin{bmatrix} x_1' H^{-1} & x_1' H^{-2} & \dots & x_1' H^{-t+1} \\ 0 & x_2' H^{-1} & \dots & x_2' H^{-t+2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & x_{t-1}' H^{-1} \\ 0 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} v_2 \\ v_3 \\ \vdots \\ v_t \end{bmatrix} \\ (\equiv y_t^*) \quad (\equiv \tilde{x}_t) \quad (\equiv A_t) \end{aligned} \quad (67a)$$

La matrice des variances-covariances des erreurs de (67a) est donnée par:

$$\Omega_t = \sigma^2 (I_t + A_t (I_{t-1} \otimes P) A_t') \quad (67b)$$

où $P \equiv (1/\sigma^2)V$. SANT a montré que l'estimateur des m.c.g. de β_t pour le modèle ci-dessus correspond à calculer $\beta_{t|t}$ par application récursive du filtre de Kalman. En plus, si on calcule la matrice correspondante de

¹⁵Plus précisément, SANT a montré pour $H = I$ mais le passage à H non-singulière est direct (CHOW (1981, 1983, 1984)). Les résultats de SANT sont une généralisation de ceux de DUNCAN et HORN (1972) qui exigeaient la non-singularité de V .

variances-covariances, on obtient $\Sigma_{t|t}$ ¹⁶.

De façon analogue à (66),

$$\beta_{t+s} = H^s \beta_t + H^{s-1} v_{t+1} + \dots + H v_{t+s-1} + v_{t+s} \quad (68)$$

et, par substitution dans (54a):

$$\begin{bmatrix} y_{t+1} \\ y_{t+2} \\ \vdots \\ y_{t+n} \end{bmatrix} = \begin{bmatrix} x'_{t+1} H \\ x'_{t+2} H^2 \\ \vdots \\ x'_{t+n} H^n \end{bmatrix} \beta_t + \left\{ \begin{bmatrix} u_{t+1} \\ u_{t+2} \\ \vdots \\ u_{t+n} \end{bmatrix} - \begin{bmatrix} x'_{t+1} & 0 & \dots & 0 \\ x'_{t+2} H & x'_{t+2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ x'_{t+n} H^{n-1} & x'_{t+n} H^{n-2} & \dots & x'_{t+n} \end{bmatrix} \begin{bmatrix} v_{t+1} \\ v_{t+2} \\ \vdots \\ v_{t+n} \end{bmatrix} \right\} \quad (69a)$$

($\equiv B_t$)

La matrice des variances-covariances des erreurs de (69a) est donnée par:

$$\sigma^2 (I_n + B_t (I_n \otimes P) B_t') \quad (69b)$$

SANT a aussi montré que si on applique les m.c.g. à l'ensemble (67)-(69) on obtiendra le lissage $\beta_{t|t+n}$.

Une remarque importante: l'approche de SANT permet d'évaluer $\beta_{t|t}$ et $\Sigma_{t|t}$ uniquement pour $t \geq K$ (parce que pour $t < K$ on n'a pas assez de degrés de liberté pour estimer (67)); mais, par contre, elle n'exige pas l'établissement préalable des conditions initiales $\beta_{0|0}$ et $\Sigma_{0|0}$. Au fait, l'estimation de (67) restreinte aux K premières observations permet de calculer

$$\beta_{K|K} = (x'_t \Omega_t^{-1} x_t)^{-1} x'_t \Omega_t^{-1} y_t^* = x_t^{-1} y_t^* \quad (70a)$$

$$\Sigma_{K|K} = (x'_t \Omega_t^{-1} x_t)^{-1} \quad (70b)$$

qui peuvent être utilisées comme conditions initiales de la procédure

¹⁶Connaissant $\beta_{t|t}$ et $\Sigma_{t|t}$, (56) et (60) donnent $\beta_{t+1|t}$ et $\Sigma_{t+1|t}$.

itérative définie par les expressions (56) à (61).

Considérons maintenant l'estimation de σ^2 et des paramètres inconnus de H et de P. Les différences

$$y_t - y_{t|t-1} = y_t - x_t' \beta_{t|t-1} \quad (t=1, \dots, T) \quad (71)$$

sont indépendantes et normalement distribuées avec:

$$E(y_t - x_t' \beta_{t|t-1} | \mathcal{I}_{t-1}) = 0$$

$$V(y_t - x_t' \beta_{t|t-1} | \mathcal{I}_{t-1}) = \sigma^2 + x_t' \Sigma_{t|t-1} x_t$$

où $\beta_{t|t-1}$ et $\Sigma_{t|t-1}$ dépendent des paramètres inconnus de H et P. Notons θ le vecteur formé par ces paramètres inconnus. La log-vraisemblance de $(y_{K+1}, \dots, y_T)$ conditionnelle à $(y_1, \dots, y_K)$ permet d'estimer σ^2 et θ :

$$\ln L = \text{cte} - \frac{1}{2} \sum_{t=K+1}^T \ln(x_t' \Sigma_{t|t-1} x_t + \sigma^2) - \frac{1}{2} \sum_{t=K+1}^T \frac{(y_t - x_t' \beta_{t|t-1})^2}{x_t' \Sigma_{t|t-1} x_t + \sigma^2} \quad (72)$$

Pour θ et σ^2 donnés, on pourra évaluer $\beta_{t|t-1}$ et $\Sigma_{t|t-1}$ par une des deux procédures précédemment décrites et substituer après dans (72), de façon à calculer $\ln L$ pour les valeurs en question. La maximisation de la vraisemblance par rapport à θ et σ^2 se fait en utilisant des méthodes numériques¹⁷. PAGAN (1980) discute des conditions d'identification de θ et de σ^2 et les propriétés asymptotiques de ces estimateurs du maximum de vraisemblance.

ROSENBERG (1973), HARVEY (1978) et HARVEY et PHILLIPS (1982), parmi d'autres auteurs, ont étudié un cas particulier important de la formulation (54), connu dans la littérature sous le nom de modèle du retour à la

¹⁷CHOW (1981, 1983, 1984) résume deux algorithmes de maximisation possibles.

normalité (Return to Normality Model):

$$y_t = x_t' \alpha_t + u_t \quad (73a)$$

$$(\alpha_t - \bar{\alpha}) = R(\alpha_{t-1} - \bar{\alpha}) + \eta_t \quad (t=1, \dots, T) \quad (73b)$$

où R ($K \times K$) est une matrice avec toutes les valeurs propres inférieures à 1 en module, $u_t \sim \text{n.i.d.}(0, \sigma^2)$ et $\eta_t \sim \text{n.i.d.}(0, \Phi)$. La considération du processus (73b) suppose que les coefficients évoluent stochastiquement dans le temps, mais montrent une tendance à converger vers $\bar{\alpha}$ ("la normalité")¹⁸.

La formulation (73) peut être envisagée comme un cas particulier de (54):

$$y_t = [x_t' \ x_t'] \begin{bmatrix} \bar{\alpha}_t \\ \delta_t \end{bmatrix} + u_t \quad (74a)$$

$$\begin{bmatrix} \bar{\alpha}_t \\ \delta_t \end{bmatrix} = \begin{bmatrix} I_K & 0 \\ 0 & R \end{bmatrix} \begin{bmatrix} \bar{\alpha}_{t-1} \\ \delta_{t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ \eta_t \end{bmatrix} \quad (74b)$$

où $\delta_t = \alpha_t - \bar{\alpha}$. Les méthodes d'estimation discutées pour le cas général peuvent être utilisées pour estimer $\beta_t' = [\bar{\alpha}' \ \delta_t']$ (qui est maintenant de dimension $2K$). Cependant, les particularités de cette formulation stationnaire ont permis d'introduire quelques modifications à la procédure d'estimation au niveau de l'obtention des conditions de départ. Si on calcule ces conditions initiales par (70) on utilise $2K$ observations. Ce nombre est réduit à la moitié si on suppose que le système part d'une situation d'équilibre stochastique stationnaire. Remarquons que cela change

¹⁸SWAMY et TINSLEY (1980) généralisent la formulation (73) substituant $\bar{\alpha}$ par $\bar{B}z_t$, où $\bar{B}$ est une matrice de constantes et z_t est un vecteur de variables exogènes.

toute la logique de l'estimation: la log-vraisemblance (72) est conditionnelle à $(y_1, \dots, y_K)$ et la vraisemblance basée sur l'hypothèse de stationarité ne le sera plus. Ecrivons (74a) pour les K premières observations:

$$\begin{aligned} y_t &= x_t' \bar{\alpha} + \varepsilon_t \quad (t=1, \dots, K) \\ \Leftrightarrow y_K^* &= X_K^* \bar{\alpha} + \varepsilon_K^* \end{aligned} \quad (75)$$

où $\varepsilon_t = x_t' \delta_t + u_t$, $y_K^* = [y_1 \dots y_K]$, $X_K^* = [x_1 \dots x_K]$ et $\varepsilon_K^* = [\varepsilon_1 \dots \varepsilon_K]$.

L'hypothèse d'équilibre stochastique stationnaire consiste à admettre que (pour $t=1, \dots, K$):

$$\begin{cases} E(\delta_t) = 0 & 19 \\ E(\delta_t \delta_t') = \Gamma \end{cases}$$

Sous cette hypothèse, la matrice des variances-covariances des erreurs de (75)

$$\begin{aligned} \varepsilon_K^* &= \begin{bmatrix} x_1' & 0 \\ & x_2' & \\ & & \ddots & \\ 0 & & & x_K' \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_K \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_K \end{bmatrix} \\ &= \tilde{X}_K (\delta_K^*) (\equiv u_K^*) \end{aligned} \quad (77)$$

est donnée par:

$$W_K = \tilde{X}_K E(\delta_K^* \delta_K^{*'}) \tilde{X}_K' + \sigma^2 I_K \quad (78)$$

avec $E(\delta_t \delta_t') = \Gamma$ et $E(\delta_t \delta_s') = R^{t-s} \Gamma$. Par application des m.c.g. à (75) on

¹⁹ Γ peut être exprimée en fonction de R et de Φ par résolution du système:

$$\Gamma = V(\delta_t) = V(R\delta_{t-1} + \eta_t) = R\Gamma R' + \Phi \Leftrightarrow [I - (R\otimes R)]\text{vec}(\Gamma) = \text{vec}(\Phi)$$

Dans le cas où R est diagonale, la résolution du système se simplifie: $\gamma_{ij} = \phi_{ij} / (1 - r_{ii} r_{jj})$.

obtient:

$$\bar{\alpha}_{k|k} = (X_k^* W_k^{-1} X_k^*)^{-1} X_k^* W_k^{-1} y_k^* = (X_k^*)^{-1} y_k^* \quad (79a)$$

$$V(\bar{\alpha}_{k|k}) = (X_k^*)^{-1} W_k (X_k^*)^{-1} \quad (79b)$$

d'où suivent directement les conditions initiales pour l'estimation de (74):

$$\beta_{k|k} = \begin{bmatrix} \bar{\alpha}_{k|k} \\ 0 \end{bmatrix} \quad (80a)$$

$$\Sigma_{k|k} = \begin{bmatrix} V(\bar{\alpha}_{k|k}) & E[(\bar{\alpha}_{k|k} - \bar{\alpha}) \delta_k^{*'}] \\ \dots & \Gamma \end{bmatrix} \quad (80b)$$

avec $E[(\bar{\alpha}_{k|k} - \bar{\alpha}) \delta_k^{*'}] = E[(X_k^*)^{-1} c_k^* \delta_k^{*'}] = (X_k^*)^{-1} \tilde{X}_k E(\delta_k^* \delta_k^{*'})$.

Pour terminer, une remarque sur le cas de données mixtes chronologiques/transversales, considéré originalement par ROSENBERG (1973). Dans la formulation pour données mixtes, les coefficients sont supposés identiques pour tous les individus et évoluent uniquement dans le temps:

$$\begin{cases} y_t = X_t \alpha_t + u_t \\ (N \times 1) \quad (N \times K) \quad (K \times 1) \quad (N \times 1) \end{cases} \quad (81a)$$

$$\begin{cases} (\alpha_t - \bar{\alpha}) = R(\alpha_{t-1} - \bar{\alpha}) + \eta_t \end{cases} \quad (t=1, \dots, T) \quad (81b)$$

$$\begin{cases} y_t = [X_t \quad X_t] \begin{bmatrix} \bar{\alpha}_t \\ \delta_t \end{bmatrix} + u_t \\ \begin{bmatrix} \bar{\alpha}_t \\ \delta_t \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & R \end{bmatrix} \begin{bmatrix} \bar{\alpha}_{t-1} \\ \delta_{t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ \eta_t \end{bmatrix} \end{cases}$$

avec $u_t \sim \text{n.i.d.}(0, \sigma^2 Q)$. Supposons d'abord que la matrice Q est connue. Pour $N \geq K$, les N observations transversales de l'époque 1 peuvent être utilisées pour établir les conditions de départ de la procédure récursive (56)-(61), qui continue à être valable à quelques détails près. Les conditions initiales peuvent être établies supposant que $t = 1$ correspond à

un équilibre stationnaire stochastique, c'est-à-dire, que $E(\delta_1) = 0$ et que $V(\delta_1) = \Gamma$. Si la matrice Q n'est pas connue: 1) on estimera les erreurs $\hat{u}_t = y_t - x_t'(\hat{\alpha} + \delta_{t|T})$ ($t=1, \dots, T$) avec $Q = I$; 2) on estimera après Q utilisant les erreurs $\hat{u}_t$; 3) on réestimera ensuite le modèle avec $\hat{Q}$. La procédure pourra être poursuivie jusqu'à convergence.

REFERENCES

- AMEMIYA, T. (1977) "A Note on a Heteroscedastic Model", *Journal of Econometrics*, 6, pp.365-370.
- AMEMIYA, T. (1978) "A Note on a Random Coefficients Model", *International Economic Review*, 19, pp.793-796.
- AMEMIYA, T. (1985) *Advanced Econometrics*, Cambridge, Mass.: Harvard Univ. Press.
- BELSLEY, D. A. (1973) "On the Determination of Systematic Parameter Variation in the Linear Regression Model", *Annals of Economic and Social Measurement*, 2/4, pp.487-494.
- BREUSCH, T. S.; PAGAN, A. R. (1979) "A Simple Test for Heteroscedasticity and Random Coefficient Variation", *Econometrica*, 47, pp.1287-1294.
- CHOW, G. C. (1981) *Econometric Analysis by Control Methods*, New York: John Wiley & Sons.
- CHOW, G. C. (1983) *Econometrics*, New York: McGraw-Hill.
- CHOW, G. C. (1984) "Random and Changing Coefficient Models" in (Eds.) GRILICHES, Z.; INTRILIGATOR, M. D. *Handbook of Econometrics*, vol II, chap. 21, Amsterdam: North-Holland.
- COOLEY, T. F.; PRESCOTT, E. C. (1973a) "Systematic (Non-Random) Variation Models - Varying Parameter Regression: A Theory and Some Applications", *Annals of Economic and Social Measurement*, 2/4, pp.463-473.
- COOLEY, T. F.; PRESCOTT, E. C. (1973b) "An Adaptive Regression Model", *International Economic Review*, 14, pp.364-371.
- COOLEY, T. F.; PRESCOTT, E. C. (1973c) "Tests of an Adaptive Regression Model", *Review of Economics and Statistics*, 55, pp.248-256.
- COOLEY, T. F.; PRESCOTT, E. C. (1976) "Estimation in the Presence of Stochastic Parameter Variation", *Econometrica*, 44, pp.167-184.
- EUBANK, R. L. (1988) *Spline Smoothing and Nonparametric Regression*, New York: Marcel Dekker.

- FISK, P. R. (1967) "Models of the Second Kind in Regression Analysis", *Journal of the Royal Statistical Society, Series B*, 29, pp.266-281.
- FOMBY, T. B.; HILL, R. C.; JOHNSON, S. R. (1984) *Advanced Econometric Methods*, New York: Springer.
- FROELICH, B. R. (1973) "Some Estimators for a Random Coefficient Regression Model", *Journal of the American Statistical Association*, 68, pp. 329-335.
- GOLDFELD, S. M.; QUANDT, R. E. (1972) *Nonlinear Methods in Econometrics*, Amsterdam: North-holland.
- GRIFFITHS, W. F. (1972) "Estimating Actual Response Coefficients in the Hildreth-Houck Random Coefficient Model", *Journal of the American Statistical Association*, 67, pp.633-635.
- HARDLE, W. (1990) *Applied Nonparametric Regression*, Econometric Society Monographs, Cambridge: Cambridge Univ. Press.
- HARVEY, A. C. (1978) "The Estimation of Time-Varying Parameters From Panel Data", *Annales de l'INSEE*, 30-31, pp.203-226.
- HARVEY, A. C.; PHILLIPS, G. D. A. (1982) "The Estimation of Regression Models with Time-Varying Parameters", in (Eds.) DEISTLER, M.; FÜRST, E.; SCHWÖDIAUER, G., *Games, Economic Dynamics, and Time Series Analysis*, Cambridge, Mass.: Physica-Verlag.
- HILDRETH, C.; HOUCK, J. P. (1968) "Some Estimators for a Linear Model with Random Coefficients", *Journal of the American Statistical Association*, 63, pp.584-595.
- HSIAO, C. (1974) "Statistical Inference for a Model with Both Random Cross-Sectional and Time Effects", *International Economic Review*, 15, pp.12-30.
- HSIAO, C. (1975) "Some Estimation Methods for a Random Coefficient Model", *Econometrica*, 43, pp.304-325.
- HSIAO, C. (1986) *Analysis of Panel Data*, Econometric Society Monographs, Cambridge: Cambridge Univ. Press.
- HURWICZ, L. (1950) "Systems with Nonadditive Disturbances", in (Ed.) KOOPMANS, T. C., *Statistical Inference in Dynamic Economic Models*, New York: Wiley.
- JOHNSON, L. W. (1977) "Stochastic Parameter Regression: An Annotated Bibliography", *International Statistical Review*, 45, pp.257-272.
- JOHNSON, L.W. (1980) "Stochastic Parameter Regression: Additional Annotated Bibliography", *International Statistical Review*, 48, pp.95-102.
- JUDGE, G. G.; GRIFFITHS, W. E.; HILL, R. C.; LEE, T.-C. (1980), *The Theory and Practice of Econometrics*, New York: John Wiley & Sons.
- KALMAN, R. E. (1960) "A New Approach to Linear Filtering and Prediction

- Problems", *Transactions of ASME, Series D, Journal of Basic Engineering*, 82, pp.35-45.
- KALMAN, R. E.; BUCY, R. (1961) "New Results in Linear Filtering and Prediction Theory", *Transactions of ASME, Series D, Journal of Basic Engineering*, 83, pp. 95-108.
- KLEIN, L. R. (1953), *A Textbook of Econometrics*, Evanston, Ill.: Row, Peterson and Co..
- MADDALA, G. S. (1977), *Econometrics*, New York: McGraw-Hill.
- MADDALA, G. S. (1983), *Limited Dependent and Qualitative Variables in Econometrics*, New York: Cambridge Univ. Press.
- MADDALA, G. S. (1986) "Disequilibrium, Self-Selection, and Switching Models", in (Eds.) GRILICHES, Z.; INTRILIGATOR, M. D., *Handbook of Econometrics*, vol III, Elsevier Science Publishers BV.
- PAGAN, A. (1980) "Some Identification and Estimation Results for Regression Models with Stochastically Varying Coefficients", *Journal of Econometrics*, 13, pp.341-363.
- PINHEIRO, M. (1990) "Introduction à la Régression Non-Paramétrique: Les Estimateurs Kernel", *Economia*, XIV, pp.441-466.
- QUANDT, R. E. (1982) "Econometric Disequilibrium Models" + "Comments", *Econometric Reviews*, vol 1, No 1, pp.1-96.
- RAJ, B.; ULLAH, A. (1981), *Econometrics - A Varying Coefficients Approach*, London: Croom Helm.
- RAO, R. C. (1965) "The Theory of Least Squares When Parameters Are Stochastic and Its Application to the Analysis of Growth Curves", *Biometrika*, 52, pp.447-458.
- ROSENBERG, B. (1973) "Random Coefficients Models: The Analysis of a Cross Section of Time Series by Stochastically Convergent Parameter Regression", *Annals of Economic and Social Measurement*, 2/4, pp.399-428.
- RUBIN, H. (1950) "Note on Random Coefficients", in (Ed.) KOOPMANS, T. C. *Statistical Inference in Dynamic Economic Models*, Wiley: New York.
- SAGE, A. P.; MELSA, J. L. (1971), *Estimation Theory with Applications to Communications and Control*, New York: McGraw-Hill.
- SANT, D. T. (1977) "Generalized Least Squares Applied to Time Varying Parameter Models", *Annals of Economic and Social Measurement*, 6/3, pp.301-311.
- SINGH, B.; NAGAR, A. L.; CHOUDHRY, N. K.; RAJ, B. (1976) "On the Estimation of Structural Change: a Generalisation of the Random Coefficients Regression Model", *International Economic Review*, 17, pp.340-361.
- SRIVASTAVA, V. K.; MISHRA, G. D.; CHATURVEDI, A. (1981) "Estimation of Linear Regression Model with Random Coefficients Ensuring Almost

- Non-Negativity of Variance Estimators", *Biometrical Journal*, 23, pp.3-8.
- SWAMY, P. A. V. B. (1970) "Efficient Inference in a Random Coefficient Regression Model", *Econometrica*, 38, pp.311-323.
- SWAMY, P. A. V. B. (1971), *Statistical Inference in Random Coefficient Regression Models*, New York: Springer-Verlag.
- SWAMY, P. A. V. B. (1974) "Linear Models with Random Coefficients", in (Ed.) ZAREMBKA, P., *Frontiers in Econometrics*, New York: Academic Press.
- SWAMY, P. A. V. B.; MEHTA, J. S. (1975) "Bayesian and Non-Bayesian Analysis of Switching Regressions and of Random Coefficient Regression Models", *Journal of the American Statistical Association*, 70, pp.593-602.
- SWAMY, P. A. V. B.; MEHTA, G. S. (1977) "Estimation of Linear Models with Time and Cross-Sectionally Varying Coefficients", *Journal of the American Statistical Association*, 72, pp.890-898.
- SWAMY, P. A. V. B.; TINSLEY, P. A. (1980) "Linear Prediction and Estimation Methods for Regression Models with Stationary Stochastic Coefficients", *Journal of Econometrics*, 12, pp.103-142.
- THEIL, H. (1961) *Economic Forecasts and Policy*, 2nd ed., Amsterdam: North-Holland.
- THEIL, H.; MENNES, L. B. M. (1959) "Conception Stochastique de Coefficients Multiplicateurs Dans l'Ajustement Linéaire de Séries Temporelles", *Publications de l'Institut de Statistique de l'Univ. de Paris*, 8, pp.211-227.
- VILARES, M. J. (1986) *Structural Change in Macroeconomic Models - Theory and Estimation*, Dordrecht: Martinus Nijhoff.
- WACHTER, M. L. (1970) "Relative Wage Equations for U.S. Manufacturing Industries 1947-1967", *Review of Economics and Statistics*, 52, pp.405-410.
- WALD, A. (1947) "A Note on Regression Analysis", *Annals of Mathematical Statistics*, 18, pp.586-589.
- WHITE, H. (1980) "Using Least Squares to Approximate Unknown Regression Functions", *International Economic Review*, 21, pp.149-170.
- ZELLNER, A. (1966) "On the Aggregation Problem: a New Approach to a Troublesome Problem", Report 6628, Center For Mathematical Studies in Business and Economics, Univ. Chicago, publié dans (Eds.) FOX et al (1969) *Economic Models, Estimation and Risk Programming: Essays in Honour of Gerhard Tintner*, New York: Springer-Verlag.