

Guy MÉLARD

Analyse des données chronologiques

Séminaire de mathématiques supérieures
- Séminaire scientifique OTAN (NATO
Advanced Study Institute), Volume 89.
Édition 1985, Réédition 2012

Les Presses de l'Université de Montréal

*Je dédie cette monographie à mon collègue et
ami Roch Roy qui m'a suggéré de la réaliser
pour la première fois en 1982.*

Avant-propos

Cette monographie a été publiée pour la première fois en 1985 par les Presses de l'Université de Montréal (C.P. 6128, succ. "A", Montréal (Québec), Canada, H3C 3J7). Elle avait été réalisée dans la collection Séminaire de mathématique supérieures, Séminaire scientifique OTAN (NATO Advanced Study Institute) par le Département de mathématiques et de statistique - Université de Montréal. Il s'agissait de mes notes de cours à la vingt-et-unième session du Séminaire de mathématiques supérieures/Séminaire scientifique OTAN (ASI 82/62), tenue au Département de mathématiques et de statistique de l'Université de Montréal du 26 juillet au 13 août 1982. Cette session avait pour titre général "Analyse des données" et était placée sous les auspices de l'Organisation du Traité de l'Atlantique Nord, du ministère de l'Éducation du Québec, du Conseil de recherches en sciences naturelles et en génie du Canada et de l'Université de Montréal.

Je remercie Monsieur Alain-Nicolas RENAUD, Éditeur des Presses de l'Université de Montréal, de m'avoir autorisé de publier à nouveau ce texte de 1985 épuisé dans sa parution d'origine ainsi que d'employer le matériel dans un nouveau livre en cours d'élaboration en collaboration avec ma collègue Rajae AZRAK. Je remercie celle-ci pour la composition en \LaTeX d'une grande partie du texte. Ce travail est rendu possible par un crédit aux chercheurs du Fonds de la Recherche Scientifique. Je remercie le centre ECARES de l'Université libre de Bruxelles qui m'accueille, tous mes collègues et anciens collaborateurs, la Solvay Brussels School of Economics and Management qui me permet de poursuivre mon enseignement et Madame Jacqueline BOTTEMANNE -DOUILLY, anciennement de l'Institut de Statistique, maintenant Département de Mathématique, qui a dactylographié la plus grande partie de mes anciens articles cités ici.

La préface est en fait l'introduction de l'édition de 1985. A part la nécessité de recomposer l'ouvrage, peu de modifications ont été apportées : quelques fautes de frappe et des numéros d'équation ou de proposition en double ont été corrigées. Les références de quelques articles à paraître ont été mises à jour. Notons que le contenu du chapitre 5 a été amplifié dans le chapitre 10 de MÉLARD, G. (1990), *Méthodes de prévision à court terme*, Éditions de l'Université de Bruxelles, Bruxelles, et

Éditions Éllipses, Paris, réédité en 2007. Une version plus élaborée du chapitre 7 a été publiée dans AZRAK, R. et MÉLARD, G. (1993), Exact maximum likelihood estimation for extended ARIMA models”, in T. Subba Rao (Ed.), *Developments in Time Series, in honor of Maurice B. Priestley*, Chapman and Hall, London, pp. 110-123. Le logiciel Time Series Expert permet de traiter ces modèles. Enfin, ajoutons que la théorie asymptotique des modèles ARMA à coefficients dépendant du temps a été élaborée depuis, voir AZRAK, R., MÉLARD, G. (2006), Asymptotic properties of quasi-maximum likelihood estimators for ARMA models with time-dependent coefficients, *Statistical Inference for Stochastic Processes* 9, 2006, 279 - 330.

Bruxelles, avril 2012

Guy Mélard

ECARES CP 114/4, Université libre de Bruxelles, Avenue Franklin Roosevelt 50,
B-1050 Bruxelles
Adresse de courrier électronique : gmelard@ulb.ac.be

Préface

Ceci est l'introduction du texte original de 1985.

INTRODUCTION

La littérature sur l'analyse des données chronologiques est maintenant très vaste. Puisqu'il existe d'excellentes synthèses, nous préférons développer dans ces notes une approche qui ne fait souvent l'objet que d'un traitement marginal : la représentation géométrique des processus aléatoires du second ordre, et en particulier des processus ARMA, dans un espace de Hilbert. Bien que certains résultats ne peuvent être établis aisément que par une autre méthode (théorie des équations aux différences, théorie des fonctions analytiques, analyse spectrale), la méthode géométrique a l'avantage d'être simple et de maintenir longtemps l'exposé à un niveau général. Cela permet de traiter les processus ARMA évolutifs et de ne passer aux processus stationnaires qu'au moment approprié. On montrera plusieurs applications théoriques et pratiques de cette approche géométrique.

Deux autres aspects ont été privilégiés : les problèmes liés à l'estimation des paramètres d'un modèle et les algorithmes correspondants et une classe de modèles ARIMA généralisés qui permettent de représenter des données chronologiques dans des conditions moins restrictives que celles qui prévalent généralement.

Cette monographie est subdivisée en sept chapitres. Le chapitre 1 a pour objet de mieux situer l'approche géométrique dans le cadre de l'analyse des séries chronologiques. Le chapitre 2 présente les définitions de base relatives aux processus aléatoires du second ordre. Les chapitres 3 et 4 sont consacrés aux processus ARMA à coefficients dépendant du temps, appelés processus ARMA évolutifs, et aux processus ARMA à coefficients constants, respectivement. Le chapitre 5 reprend brièvement la méthode d'analyse des données chronologiques par les modèles ARIMA selon la méthode de Box et Jenkins. Ceci permet de mieux situer les deux chapitres suivants dans le cadre général. Le chapitre 6 est consacré à la méthode d'estimation des paramètres par application du principe du maximum de vraisemblance. Le chapitre 7 vise à étendre la classe de modèles envisagée au chapitre 5.

La bibliographie ne prétend pas être complète. Une bibliographie récente peut être trouvée dans Newbold (1981).

Je tiens à exprimer ma profonde reconnaissance aux organisateurs qui n'ont pas ménagé leur peine pour que ce séminaire soit un succès : M. Pierre Berthiaume, Directeur du séminaire et éditeur, les directeurs scientifiques MM. Yves Lepage et Carl E. Särndal ainsi que la secrétaire Mlle Ghislaine David. J'adresse aussi mes remerciements à mes collègues et amis Roch Roy, Piero Barone, Marc Hallin et Jacques Beaudry qui m'ont communiqué leurs remarques. J'ai une pensée pour Mademoiselle Johanne Beausoleil qui a transformé

mon manuscrit brouillon en un texte qui a la rigueur d'un poème et, je l'espère, en possède la chaleur. Quoi qu'il en soit je remercie ma muse.

Table des matières

1	APPROCHES ET MÉTHODES DE L'ANALYSE DES SÉRIES CHRONOLOGIQUES	1
	1.1. Série chronologique	1
	1.2. Réalisation d'un processus aléatoire	1
	1.3. Les différentes approches	2
	1.4. Un exemple en gestion des stocks	2
	1.5. Formulation générale	3
	1.6. Concept d'innovation et modèle multivarié	4
2	PROCESSUS ALÉATOIRES DU SECOND ORDRE	7
	2.1. Espace linéaire de variables aléatoires	7
	2.2. Espace linéaire des variables aléatoires du second ordre	7
	2.3. Espace d'Hilbert des variables aléatoires du second ordre	8
	2.4. Variables aléatoires vectorielles du second ordre	9
	2.5. Processus aléatoires	11
	2.6. Processus aléatoires du second ordre	12
3	PROCESSUS ARMA ÉVOLUTIFS	17
	3.1. Définitions et notations	18
	3.2. Condition nécessaire pour un processus purement indéterminable ...	19
	3.3. Caractérisation des processus AR et MA	22
	3.4. Inversibilité	23
	3.5. Exemples de processus ARMA évolutifs	25
	3.6. La fonction de covariance	33
	3.7. La prévision	34
4	PROCESSUS ARMA CONSTANTS	37
	4.1. Définitions	37
	4.2. Stationnarité et inversibilité	38
	4.3. Caractérisation des processus ARMA	47
	4.4. Exemples de processus ARMA	49

4.5. La fonction d'autocovariance	50
4.6. Les processus ARIMA	57
5 IDENTIFICATION, ESTIMATION ET DIAGNOSTICS DANS LES MODÈLES DE SÉRIES CHRONOLOGIQUES	59
5.1. Familiarisation avec les données	59
5.2. Analyse préliminaire	60
5.3. Identification du modèle	61
5.4. L'estimation des paramètres	62
5.5. Tests d'adéquation du modèle	63
5.6. La prévision	64
5.7. Interprétation du modèle	65
6 L'ESTIMATION DES PARAMÈTRES DE MODÈLES ARMA	67
6.1. Les méthodes basées sur la somme des carré	67
6.2. La méthode de Ljung et Box	70
6.3. La méthode de factorisation de Cholesky	72
6.4. La méthode basée sur le filtre de Kalman	74
6.5. La méthode rapide pour les modèles ARMA constants scalaires	80
6.6. Comparaison des méthodes d'évaluation de la fonction de vraisemblance	86
7 LE MODÈLE ARIMA GÉNÉRALISÉ	89
7.1. Le modèle ARIMA généralisé	89
7.2. Les interventions	91
7.3. Les transformations	93
7.4. La tendance multiplicative et la tendance en écart type des innovations	94
7.5. Le sous-modèle ARMA évolutif	94
7.6. La différence ordinaire et la tendance déterministe	95
7.7. La différence saisonnière et la composante saisonnière déterministe .	96
7.8. L'estimation des paramètres: la méthode conditionnelle	97
7.9. L'estimation des paramètres: la méthode non conditionnelle	99
Bibliographie	103
Glossaire	109

Chapitre 1

APPROCHES ET MÉTHODES DE L'ANALYSE DES SÉRIES CHRONOLOGIQUES

Dans ce chapitre on introduit le sujet au moyen d'une présentation heuristique mais assez générale de l'analyse des données chronologiques par le biais du modèle de prévision.

1.1. Série chronologique

Le *temps* est une variable dont les valeurs appartiennent à un ensemble totalement ordonné T , qui sera presque toujours l'ensemble \mathbb{Z} des entiers. On étudie une ou plusieurs variables à valeurs réelles qui évoluent en fonction du temps. On distingue une *série chronologique* (s.c., *time series*) *univariée* relative à une variable scalaire z à valeurs dans \mathbb{R} , l'ensemble des réels, et une *série chronologique multivariée* relative à une variable vectorielle \mathbf{z} , à valeurs dans \mathbb{R}^k , l'espace euclidien à k dimensions.

Les mesures de la variable se font souvent à des temps équidistants, séparés par la période d'observation qu'on peut utiliser comme unité de temps. Les observations sont relatives à un intervalle de temps appelé *intervalle d'observation* qu'on peut prendre comme étant l'ensemble $\{1, 2, \dots, n\}$; n est alors la *longueur de la s.c.*

Remarque. Nous n'aborderons pas le cas où le niveau de mesure des variables est dichotomique, nominal ou ordinal, ni celui où les observations ne sont pas équidistantes.

1.2. Réalisation d'un processus aléatoire

Une s.c. $(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n)$ est souvent considérée comme une réalisation particulière d'un *processus aléatoire* (p.a.) (voir chapitre 2). On dispose généralement d'une seule réalisation de ce p.a. Pour des raisons qui apparaissent plus loin, on suppose

souvent que ce p.a. est gaussien, ce qui implique que la distribution marginale de la *variable aléatoire* (v.a.) \mathbf{z}_t est normale.

1.3. Les différentes approches

Le problème étudié varie selon le type de données : longueur n de la série, dimension k de la variable, présence de composantes périodiques, présence d'erreurs de mesure constituant le bruit. Ainsi, en économétrie où n est petit et k est grand, on étudie les relations entre variables qui sont suggérées par la théorie économique. En management, n est plus grand et k est plus petit : on profite donc davantage de l'information contenue dans le passé des séries afin d'élaborer des prévisions pour l'avenir. La décomposition en tendance, cycle conjoncturel, saisonnier et aléa est souvent exploitée. En physique, le signal étudié comporte souvent des périodicités. Pour l'isoler du bruit on emploie le lissage et le filtrage. Une étude des périodicités s'effectue au moyen de l'approche fréquentielle ou spectrale. Étant donné la grande taille de n , les aspects algorithmiques sont importants.

La plus grande partie de l'analyse statistique des s.c. est néanmoins basée sur une approche plutôt temporelle où la fonction d'autocorrélation, en dépit de ses défauts, joue un rôle central.

1.4. Un exemple en gestion des stocks

A titre d'exemple, considérons la demande d'un article déterminé appartenant au stock d'une entreprise de vente en gros. Cette demande est exprimée en nombre d'unités demandées c'est-à-dire vendues ou refusées à la vente pour cause de pénurie. Même si le stock est important il sera possible de conserver un historique de quelques années sous forme de données mensuelles. Le but est notamment de prévoir la distribution de la demande pendant l'intervalle de temps allant jusqu'à la prochaine livraison, afin de déterminer la probabilité de pénurie. Le problème inverse de détermination du stock de sécurité est plus intéressant mais ne sera pas traité ici. Supposons, pour simplifier, que cet intervalle de temps soit de ℓ mois, ℓ entier. La demande pendant ces ℓ mois, $z[\ell]$, sera considérée comme une v.a. qui s'exprime comme la somme des demandes de ces ℓ mois. Même en supposant une loi statistique pour ces demandes mensuelles, la détermination de la distribution de $z[\ell]$ n'est pas simple sauf si l'on suppose, par exemple, que la distribution liée des demandes mensuelles futures est normale, avec pour moyenne la somme des prévisions mensuelles, dont la prévision totale.

1.5. Formulation générale

La présentation heuristique que nous développons maintenant a pour but d'introduire l'approche temporelle par modèles. En effet, la plupart des modèles qu'on rencontre dans la littérature sont des cas particuliers de cette formulation générale. Considérons une variable d'intérêt \mathbf{z} étudiée en fonction du temps.

Soit \mathbf{y} une *variable explicative* de \mathbf{z} , c'est-à-dire une variable dont les valeurs vont servir à prévoir \mathbf{z} . Notons \mathcal{I}_t l'information en t , constituée des valeurs passées de z et de y

$$\mathcal{I}_t = (\mathbf{z}_{t-i}; i \geq 1) \oplus (\mathbf{y}_{t-i}; i \geq 1).$$

Le symbole \oplus utilisé ici recevra une signification précise au chapitre 2. Le fait que l'information la plus récente disponible sur \mathbf{z} et sur \mathbf{y} soit contemporaine est sans importance. Dans le cas contraire où l'on connaîtrait \mathbf{z} avec b unités de temps de retard, on peut noter $\mathbf{z}_t^* = \mathbf{z}_{t-b}$.

Supposons qu'on s'intéresse au présent et au futur de \mathbf{z} , noté \mathcal{F}_t , avec un certain horizon ℓ , d'où

$$\mathcal{F}_t = (\mathbf{z}_{t+i}; 0 \leq i \leq \ell).$$

On devrait rechercher la distribution liée de \mathcal{F}_t étant donné l'information disponible, c'est-à-dire la *fonction de distribution* (f.d.) $F(\mathcal{F}_t|\mathcal{I}_t)$. On se limite souvent à la moyenne de \mathcal{F}_t soit

$$\bar{\mathcal{F}}_t = \int \mathcal{F}_t dF(\mathcal{F}_t|\mathcal{I}_t).$$

Le problème est de choisir cette f.d. C'est ce qu'on appelle l'élaboration d'un modèle de prévision. Cela ne peut se faire que moyennant des hypothèses parmi les suivantes qui ont pour but de réduire la classe des f.d. candidates.

Hypothèse H₁ ou hypothèse de normalité : $F(\mathcal{F}_t|\mathcal{I}_t)$ est une f.d. normale;

Hypothèse H₂ ou hypothèse de linéarité : $\bar{\mathcal{F}}_t$ est une fonction linéaire de \mathcal{I}_t :
 $\bar{\mathcal{F}}_t = \varphi(\mathcal{I}_t)$;

Hypothèse H₃ ou hypothèse de stationnarité : pour tout $h > 0$, $F(\mathcal{F}_{t+h}|\mathcal{I}_{t+h}) = F(\mathcal{F}_t|\mathcal{I}_t)$.

Remarques. Ces hypothèses ne sont pas imposées systématiquement. Elles peuvent être allégées comme suit :

1. avant d'appliquer H₁, on peut effectuer une transformation préalable sur les variables;
2. dans H₂, la linéarité par rapport à \mathbf{y}_t n'est pas toujours nécessaire;
3. l'hypothèse H₃ peut être adoucie en spécifiant la manière avec laquelle la f.d. évolue en fonction de t .

Sous l'hypothèse H₂

$$\bar{\mathcal{F}}_t = \sum_{i=1}^{\infty} \pi'_i \mathbf{z}_{t-i} + \sum_{i=1}^{\infty} \pi''_i \mathbf{y}_{t-i},$$

où Π'_i et Π''_i sont des matrices à coefficients constants grâce à H_3 . En utilisant H_1 , on a

$$\mathcal{F}_t = \sum_{i=1}^{\infty} \Pi'_i \mathbf{z}_{t-i} + \sum_{i=1}^{\infty} \Pi''_i \mathbf{y}_{t-i} + E_t, \quad (1.1)$$

où $E_t \sim N(0; \Sigma)$ (ce qui représente la loi normale de vecteur moyenne 0 et de matrice covariance Σ). Notons que Σ est une matrice constante par rapport à t , grâce à H_3 .

1.6. Concept d'innovation et modèle multivarié

En isolant \mathbf{z}_t dans F_t , il vient de (1.1)

$$\mathbf{z}_t = \sum_{i=1}^{\infty} \pi'_i \mathbf{z}_{t-i} + \sum_{i=1}^{\infty} \pi''_i \mathbf{y}_{t-i} + \mathbf{e}_t, \quad (1.2)$$

où $\mathbf{e}_t \sim N(0; \Sigma)$ et les matrices π'_i , π''_i et Σ sont des blocs extraits de Π'_i , Π''_i et Σ , respectivement. Les \mathbf{e}_t doivent être interprétés comme les *erreurs de prévision d'horizon 1*. On les appellera *innovations*. La justification de cette terminologie est la suivante: $\mathbf{e}_t = \mathbf{z}_t - \hat{\mathbf{z}}_t$, où $\hat{\mathbf{z}}_t$ est la prévision de \mathbf{z}_t faite à l'aide de l'information \mathcal{I}_t mais \mathbf{e}_t est aussi l'apport neuf en t puisque les autres termes de (1.2) sont connus (à supposer que les π'_i et π''_i sont connus). Au sujet des innovations on ajoute une hypothèse supplémentaire:

Hypothèse H_4 ou hypothèse d'indépendance des innovations : les \mathbf{e}_t sont des v.a. indépendantes, ce qui implique que $\forall t, s \neq t \quad : \quad \mathbf{e}_t \perp \mathbf{e}_s$ (c'est-à-dire \mathbf{e}_t et \mathbf{e}_s sont des v.a. indépendantes).

Cette hypothèse peut être allégée en une hypothèse de non autocorrélation, jusqu'à un certain point dans la théorie :

$$\forall t, s \neq t : \mathbf{e}_t \perp \mathbf{e}_s$$

c'est-à-dire les composantes $(\mathbf{e}_t)_i$ de \mathbf{e}_t et $(\mathbf{e}_s)_j$ de \mathbf{e}_s sont des v.a. *non corrélées* ($i, j = 1, 2, \dots, k$).

La forme (1.2) est similaire au *modèle de fonction de transfert*. On montrera à la section 3.7 que (1.2) permet en fait de déterminer (1.1) La distinction entre \mathbf{z} et \mathbf{y} peut être atténuée en supposant pour \mathbf{y} une décomposition similaire à (1.2)

$$\mathbf{y}_t = \sum_{i=1}^{\infty} \pi'''_i \mathbf{y}_{t-i} + \mathbf{e}'_t,$$

et en réécrivant (1.2) comme suit :

$$\begin{bmatrix} \mathbf{z}_t \\ \mathbf{y}_t \end{bmatrix} = \sum_{i=1}^{\infty} \begin{bmatrix} \pi'_i & \pi''_i \\ \mathbf{0} & \pi'''_i \end{bmatrix} \begin{bmatrix} \mathbf{z}_{t-i} \\ \mathbf{y}_{t-i} \end{bmatrix} + \begin{bmatrix} \mathbf{e}_t \\ \mathbf{e}'_t \end{bmatrix},$$

ce qui est le *modèle multivarié*. Il n'y aura donc pas de restriction à envisager

$$\mathbf{z}_t = \sum_{i=1}^{\infty} \pi_i \mathbf{z}_{t-i} + \mathbf{e}_t, \quad (1.3)$$

si ce n'est la forme particulière des matrices π_i .

Remarque. Un modèle de prévision encore plus général sur certains points a été proposé par Priestley (1980). Ce modèle s'affranchit de l'hypothèse de linéarité et utilise un ensemble d'information de dimension plus réduite grâce au concept de vecteur d'état (voir section 3.5).

Chapitre 2

PROCESSUS ALÉATOIRES DU SECOND ORDRE

Ce chapitre sert à spécifier les notations et à rappeler un certain nombre de résultats connus. En général, les démonstrations ne sont pas données. Pour la terminologie d'analyse fonctionnelle, voir par exemple Rudin (1966) ou Taylor (1958). L'approche par les espaces de Hilbert a été développée par Karhunen (1947) et Parzen (1967).

2.1. Espace linéaire de variables aléatoires

Soit (Ω, B, P) un *espace de probabilité* sur lequel seront définies toutes les variables aléatoires. Notons \mathbb{R} le champ des réels et \mathbb{C} le champ des complexes.

Définition 2.1. z est une *variable aléatoire (v.a.) réelle* ssi $z : \Omega \rightarrow \mathbb{R} : \omega \mapsto z(\omega)$ est une fonction B -mesurable.

Définition 2.2. $\mathbf{z} = (z_1, \dots, z_k)^\top$ est une *v.a. vectorielle réelle* ssi $\mathbf{z} : \Omega \mapsto \mathbb{R}^k : \omega \mapsto \mathbf{z}(\omega)$ est une fonction B -mesurable. ($^\top$ est le *symbole de transposition*.)

Notons que les z_j sont des v.a. On définit de même une v.a. complexe et une v.a. vectorielle complexe.

Proposition 2.3. *L'ensemble V de toutes les v.a. réelles (ou complexes) est un espace linéaire.*

2.2. Espace linéaire des variables aléatoires du second ordre

Définition 2.4. La *moyenne* (ou *espérance mathématique*) et le moment par rapport à l'origine d'ordre 2 de z sont définis par

$$\mu_1 = E(z) = \int_{\Omega} z(\omega)P(d\omega), \quad \mu_2 = E(|z|^2),$$

où les intégrales existent et sont finies au sens de Lebesgue, et $|\cdot|$ est la *valeur absolue* (cas réel) ou le *module* (cas complexe). L'existence de μ_2 entraîne celle de μ_1 et celle de la variance $\text{var}(z) = E(|z - \mu_1|)^2$.

Définition 2.5. z est une *v.a. du second ordre* ssi μ_2 existe.

Définition 2.6. z est une *v.a. centrée du second ordre* (v.a.s.) ssi μ_2 existe et $\mu_1 = 0$. Dans ce cas, $\text{var}(z) = \mu_2$.

Soit V_2 l'ensemble des v.a.s. appartenant à V . Si z est une variable aléatoire du second ordre de moyenne μ_1 , alors $z - \mu_1$ est une v.a.s. Dorénavant, à l'exception des chapitres 5 et 7, toutes les v.a. du second ordre seront supposées centrées.

Dans la suite, deux v.a.s. z' et z'' sont considérées comme équivalentes, ce qui est noté $z' = z''$, dès que $\text{var}(z' - z'') = 0$ et donc $P(z' = z'') = 1$.

Définition 2.7. La *covariance* entre deux v.a.s. z' et z'' , notée $\text{cov}(z', z'')$, est $E(z' z'')$ si les variables sont réelles et $E(z' \bar{z}'')$ si les variables sont complexes, où $\bar{\cdot}$ désigne le *complexe conjugué*.

Notons que $\text{cov}(z', z'') = \text{cov}(z'', z')$ pour des v.a.s. réelles tandis que $\text{cov}(z', z'') = \overline{\text{cov}(z'', z')}$ pour des v.a.s. complexes.

Proposition 2.8. (Inégalité de Schwarz). $[\text{cov}(z', z'')]^2 \leq \text{var}(z') \text{var}(z'')$. La *covariance* entre deux v.a.s. existe donc toujours.

Proposition 2.9. L'espace V_2 des v.a.s. est

1. un espace linéaire;
2. un espace normé, avec la norme $\|z\|$ telle que $\|z\|^2 = \mu_2 = \text{var}(z)$;
3. un espace métrique, avec la distance $d(z', z'') = \|z' - z''\|$

2.3. Espace de Hilbert des variables aléatoires du second ordre

Définition 2.10. Le *produit intérieur* $\langle z', z'' \rangle$ entre deux v.a.s. z' et z'' est défini par la *covariance*

$$\langle z', z'' \rangle = \text{cov}(z', z'').$$

Notons que $\langle z, z \rangle = \text{var}(z) = \|z\|^2$.

Proposition 2.11. V_2 est un espace à produit intérieur.

Proposition 2.12. V_2 est un espace topologique, avec la topologie définie à partir des boules.

Définition 2.13. Une suite de v.a.s. $z^{(1)}, \dots, z^{(n)}, \dots$ converge en moyenne vers la v.a.s. z , ce que l'on note

$$\text{l.i.m.}_{n \rightarrow \infty} z^{(n)} = z$$

(l.i.m. pour 'limit in the mean') ssi

$$\lim_{n \rightarrow \infty} \|(z^{(n)} - z)\| = 0.$$

Une suite de v.a.s. $z^{(1)}, \dots, z^{(m)}, \dots, z^{(n)}, \dots$ est une *suite de Cauchy* ssi

$$\forall \varepsilon > 0, \exists N = N(\varepsilon) : \forall n, m > N, \|z^{(n)} - z^{(m)}\| < \varepsilon.$$

Proposition 2.14. V_2 est un espace métrique complet c'est-à-dire que toute suite de Cauchy converge vers une v.a.s. V_2 est donc un espace de Hilbert.

Démonstration. Par exemple, Anderson (1971, p. 414). □

Notons S_2 un sous-espace linéaire de V_2 , fermé au sens de la convergence en moyenne, c'est-à-dire complété par les limites de suites de Cauchy convergentes. Il est clair que S_2 est un sous-espace de Hilbert de V_2 .

Définition 2.15. Deux v.a.s. z et z'' sont *non corrélées*, $z \perp z''$, ssi $\langle z', z'' \rangle = 0$. En particulier, deux v.a.s. *indépendantes* z' et z'' (ce que l'on note $z' \perp\!\!\!\perp z''$) sont non corrélées.

Définition 2.16. 1. $z \perp S_2$ ssi $\forall z' \in S_2 : z \perp z'$;
2. si S'_2 et S''_2 sont deux sous-espaces de Hilbert, $S'_2 \perp S''_2$ ssi $\forall z' \in S'_2 : z' \perp S''_2$.

Proposition 2.17. (Théorème de projection)

1. $\forall z \in V_2$: il existe un et un seul $z' \in S_2$ tel que

$$\forall z'' \in S_2 : \|z - z'\| \leq \|z - z''\|.$$

On appelle z' la projection orthogonale de z sur S_2 et on la note $P(z|S_2)$.

2. Si $z \notin S_2$, alors $[z - P(z|S_2)] \perp S_2$.

Démonstration. Par exemple, Anderson (1971, p. 418). □

Proposition 2.18. (Théorème de décomposition) $\forall z \in S_2$, il existe un et un seul $z' \in S_2$ et un et un seul $z'' \perp S_2$ tel que $z = z' + z''$.

Notons $S_2^\perp = \{z'' \in V_2 | z'' \perp S_2\}$. On écrit $V_2 = S_2 \oplus S_2^\perp$.

2.4. Variables aléatoires vectorielles du second ordre

Considérons ici \mathbb{R}^k , étant entendu que les propositions peuvent être adaptées pour \mathbb{C}^k . Notons $\mathbf{0}$ le vecteur nul ou une matrice nulle selon le contexte.

Définition 2.19. \mathbf{z} est une v.a. centrée du second ordre vectorielle (v.a.s. vectorielle) ssi $E(\mathbf{z}) = \mathbf{0}$ et $\Sigma = \int_{\Omega} \mathbf{z}\mathbf{z}^{\top} P(d\omega)$ existe et est finie. La matrice Σ est la *matrice de covariance* de \mathbf{z} . On note plus généralement le *gramien*

$$\langle \mathbf{z}', \mathbf{z}'' \rangle = \int_{\Omega} \mathbf{z}' \mathbf{z}''^{\top} P(d\omega) = \langle \mathbf{z}'', \mathbf{z}' \rangle^{\top},$$

$$\|\mathbf{z}\|^2 = \langle \mathbf{z}, \mathbf{z} \rangle = \Sigma$$

de la même façon qu'un produit intérieur. Si \mathbf{A} est une matrice $m \times k$, $\|\mathbf{A}\mathbf{z}\|^2 = \mathbf{A}\|\mathbf{z}\|^2\mathbf{A}^{\top}$. De même $\langle \mathbf{A}'\mathbf{z}', \mathbf{A}''\mathbf{z}'' \rangle = \mathbf{A}'\langle \mathbf{z}', \mathbf{z}'' \rangle\mathbf{A}''^{\top}$.

Remarque. La notation $\|\mathbf{z}\|^2$ résulte d'un abus puisqu'il s'agit d'une matrice et pas d'une norme. Nous l'avons adoptée parce qu'aucune ambiguïté ne peut en résulter.

Proposition 2.20. Si $\mathbf{z} = (z_1, \dots, z_k)^{\top}$, alors

1. \mathbf{z} est une v.a.s. vectorielle ssi z_i est une v.a.s. pour $i = 1, \dots, k$;
2. $\forall i, j = 1, \dots, k : \text{cov}(z_i, z_j) = \Sigma_{ij}$.

On note $\mathbf{z}' \perp \mathbf{z}''$ (respectivement $\mathbf{z}' \perp\!\!\!\perp \mathbf{z}''$) ssi $\forall i, j = 1, \dots, k : z'_i \perp z''_j$ (respectivement $z'_i \perp\!\!\!\perp z''_j$). Si S_2 est un sous-espace de Hilbert, on définit $P(\mathbf{z}|S_2)$ composante par composante.

Proposition 2.21. Σ est une matrice définie positive, c'est-à-dire symétrique et telle que

$$\forall \mathbf{c} \in \mathbb{R}^k : \quad \mathbf{c}^{\top} \Sigma \mathbf{c} \geq 0.$$

On note $\Sigma \geq \mathbf{0}$. De plus, si Σ est inversible, Σ est une matrice définie strictement positive, c'est-à-dire

$$\forall \mathbf{c} \in \mathbb{R}^k, \mathbf{c} \neq \mathbf{0} : \quad \mathbf{c}^{\top} \Sigma \mathbf{c} > 0,$$

ce qu'on note $\Sigma > \mathbf{0}$.

Proposition 2.22. (Factorisation de Cholesky) Si $\Sigma > \mathbf{0}$, il existe \mathbf{P} , matrice $k \times k$ triangulaire inférieure à éléments diagonaux strictement positifs telle que

$$\Sigma = \mathbf{P}\mathbf{P}^{\top}.$$

Démonstration. Par exemple, Stoer et Burlisch (1980, p.174). □

Définition 2.23. Une série de v.a.s. vectorielles $\sum_{i=1}^n \mathbf{z}^i$ converge en moyenne vers une v.a.s. vectorielle \mathbf{z} ssi

$$\lim_{n \rightarrow \infty} \left\| \sum_{i=1}^n \mathbf{z}^{(i)} - \mathbf{z} \right\| = 0.$$

Cette série converge en moyenne ssi $\forall \varepsilon > 0, \exists N = N(\varepsilon) :$

$$\forall n, m > N : \quad \left\| \sum_{i=m}^n \mathbf{z}^{(i)} \right\| < \varepsilon.$$

Proposition 2.24. Soit \mathbf{z} une v.a.s. vectorielle $k \times 1$ et S le sous-espace de Hilbert engendré par les composantes de la v.a.s. vectorielle $m \times 1$ \mathbf{v} , de matrice de covariance Σ inversible. Soit $\mathbf{C} = \langle \mathbf{z}, \mathbf{v} \rangle$. Alors

$$P(\mathbf{z}|S) = \mathbf{C}\Sigma^{-1}\mathbf{v}. \quad (2.1)$$

Démonstration. Par définition $P(\mathbf{z}|S)$ est de la forme $\mathbf{A}\mathbf{v}$ ou \mathbf{A} est une matrice $k \times m$ et $\mathbf{A}^\top = (\mathbf{a}_1, \dots, \mathbf{a}_k)$ telle que $\|z_i - \mathbf{a}_i^\top \mathbf{v}\|^2$ est minimum pour tout $i, i = 1, \dots, k$. On doit montrer que $\mathbf{A} = \mathbf{C}\Sigma^{-1}$ ou $\mathbf{A}\Sigma = \mathbf{C}$. Notons $\mathbf{C}^\top = (\mathbf{c}_1, \dots, \mathbf{c}_k)$ ou $\mathbf{c}_i^\top = \langle z_i, \mathbf{v} \rangle$. On peut écrire l'identité

$$\mathbf{a}_i^\top = \mathbf{c}_i^\top \Sigma^{-1} - (\mathbf{c}_i^\top \Sigma^{-1} - \mathbf{a}_i^\top),$$

d'où résulte

$$\begin{aligned} \|z_i - \mathbf{a}_i^\top \mathbf{v}\|^2 &= \left\| (z_i - \mathbf{c}_i^\top \Sigma^{-1} \mathbf{v}) + (\mathbf{c}_i^\top \Sigma^{-1} - \mathbf{a}_i^\top) \mathbf{v} \right\|^2 \\ &= \left\| z_i - \mathbf{c}_i^\top \Sigma^{-1} \mathbf{v} \right\|^2 + \left\| (\mathbf{c}_i^\top \Sigma^{-1} - \mathbf{a}_i^\top) \mathbf{v} \right\|^2 + 2\langle z_i - \mathbf{c}_i^\top \Sigma^{-1} \mathbf{v}, (\mathbf{c}_i^\top \Sigma^{-1} - \mathbf{a}_i^\top) \mathbf{v} \rangle \\ &= \|z_i - \mathbf{c}_i^\top \Sigma^{-1} \mathbf{v}\|^2 + (\mathbf{c}_i^\top \Sigma^{-1} - \mathbf{a}_i^\top) \Sigma (\Sigma^{-1} \mathbf{c}_i^\top - \mathbf{a}_i) \\ &\quad + 2\langle z_i, \mathbf{v} \rangle (\mathbf{c}_i^\top \Sigma^{-1} - \mathbf{a}_i) - 2\mathbf{c}_i^\top \Sigma^{-1} \langle \mathbf{v}, \mathbf{v}^\top \rangle (\Sigma^{-1} \mathbf{c}_i - \mathbf{a}_i). \end{aligned}$$

Les deux derniers termes donnent lieu à

$$2(\mathbf{c}_i^\top \Sigma^{-1} \mathbf{c}_i - \mathbf{c}_i^\top \mathbf{a}_i - \mathbf{c}_i^\top \Sigma^{-1} \mathbf{c}_i + \mathbf{c}_i^\top \mathbf{a}_i) = 0,$$

d'où

$$\|z_i - \mathbf{a}_i^\top \mathbf{v}\|^2 = \|z_i - \mathbf{c}_i^\top \Sigma^{-1} \mathbf{v}\|^2 + (\mathbf{c}_i^\top \Sigma^{-1} - \mathbf{a}_i^\top) \Sigma (\Sigma^{-1} \mathbf{c}_i - \mathbf{a}_i).$$

Dans le but d'annuler le second terme et donc de minimiser la somme, on pose $\mathbf{a}_i^\top = \mathbf{c}_i^\top \Sigma^{-1}$ c'est-à-dire $\mathbf{A} = \mathbf{C}\Sigma^{-1}$. D'où $P(\mathbf{z}|S) = \mathbf{C}\Sigma^{-1}\mathbf{v}$. \square

2.5. Processus aléatoires

Définition 2.25. Un *processus aléatoire scalaire* ou *univarié* (p.a.) $(z_t; t \in T)$ est une application $T \rightarrow V : t \mapsto z_t$. Un *p.a. vectoriel* ou *multivarié* $(\mathbf{z}_t; t \in T)$ est une application $T \rightarrow V^k : t \mapsto \mathbf{z}_t$ où $\mathbf{z}_t = (z_{t1}, \dots, z_{tk})^\top$. Notons généralement $F(\cdot)$ la fonction de distribution des v.a. (\cdot) . Un p.a. est *gaussien* ssi $\forall m > 0, \forall t_1, \dots, t_m \in T : F(\mathbf{z}_{t_1}, \dots, \mathbf{z}_{t_m})$ est une f.d. normale.

Dans les définitions suivantes, on suppose $T = \mathbb{Z}$.

Définition 2.26. $(\mathbf{z}_t; t \in \mathbb{Z})$ est un *processus strictement stationnaire* ssi $\forall j \in \mathbb{Z}, \forall m > 0, \forall t_1, \dots, t_m \in \mathbb{Z} :$

$$F(\mathbf{z}_{t_1}, \dots, \mathbf{z}_{t_m}) = F(\mathbf{z}_{t_1+j}, \dots, \mathbf{z}_{t_m+j}).$$

Définition 2.27. $(\mathbf{z}_t; t \in \mathbb{Z})$ est un *bruit blanc* (b.b.) *vectoriel* ssi $\forall m > 0, \forall t_1, \dots, t_m \in \mathbb{Z}$:

$$F(\mathbf{z}_{t_1}, \dots, \mathbf{z}_{t_m}) = F(\mathbf{z}_{t_1}) \times \dots \times F(\mathbf{z}_{t_m})$$

avec les notations de la Définition 2.25.

Définition 2.28. Deux p.a. $(\mathbf{z}'_t; t \in \mathbb{Z})$ et $(\mathbf{z}''_t; t \in \mathbb{Z})$ sont *indépendants* ssi $\forall m_1, m_2 > 0, \forall t'_1, \dots, t'_{m_1}, t''_1, \dots, t''_{m_2} \in \mathbb{Z}$:

$$F(\mathbf{z}'_{t'_1}, \dots, \mathbf{z}'_{t'_{m_1}}, \mathbf{z}''_{t''_1}, \dots, \mathbf{z}''_{t''_{m_2}}) = F(\mathbf{z}'_{t'_1}, \dots, \mathbf{z}'_{t'_{m_1}}) F(\mathbf{z}''_{t''_1}, \dots, \mathbf{z}''_{t''_{m_2}}).$$

2.6. Processus aléatoires du second ordre

Définition 2.29. Un *processus aléatoire du second ordre* (p.a.s.) $(z_t; t \in T)$ est une application $T \rightarrow V_2 : t \mapsto z_t$. Un *p.a.s. vectoriel* $(\mathbf{z}_t; t \in T)$ est une application $T \rightarrow V_2^k : t \mapsto \mathbf{z}_t$ où $\mathbf{z}_t = (z_1, \dots, z_k)^\top$.

Définition 2.30. $(\Gamma_{ts}; t, s \in T)$ est la *fonction de covariance* de $(\mathbf{z}_t; t \in T)$ ssi

$$\Gamma_{ts} = \langle \mathbf{z}_t, \mathbf{z}_s \rangle.$$

Proposition 2.31. $(\Gamma_{ts}; t, s \in T)$ est une *fonction définie positive* c'est-à-dire que tout *sous-ensemble fini* S de T :

$$\forall \mathbf{c} : S \rightarrow \mathbb{R}^k : t \mapsto \mathbf{c}_t : \sum_{t, s \in S} \mathbf{c}_t^\top \Gamma_{ts} \mathbf{c}_s \geq 0.$$

Démonstration.

$$\left\| \sum_{t \in S} \mathbf{c}_t^\top \mathbf{z}_t \right\|^2 = \sum_{t, s \in S} \mathbf{c}_t^\top \Gamma_{ts} \mathbf{c}_s \geq 0.$$

Définition 2.32. $(\mathbf{z}_t; t \in \mathbb{Z})$ est un *bruit blanc du second ordre* (b.b.s.) *vectoriel* ssi

$$\forall t, s \neq t : \mathbf{z}_t \perp \mathbf{z}_s$$

c'est-à-dire $\Gamma_{ts} = 0$ pour $t \neq s$. Notons $\Gamma_{tt} = \|\mathbf{z}_t\|^2 = \Sigma_t$, la *matrice de covariance* du b.b.s. On dit qu'un b.b.s. est un *bruit blanc du second ordre constant* (b.b.s.c.) ssi Σ_t ne dépend pas de t . On dit que c'est un *bruit blanc du second ordre normalisé* (b.b.s.n.) ssi $\Sigma_t = \mathbf{I}_k$. Un b.b.s. gaussien est un b.b. Considérons un b.b. tel que Σ_t existe et est finie. On définit alors par analogie un *bruit blanc constant* (b.b.c.) et un *bruit blanc normalisé* (b.b.n.).

Proposition 2.33. Si $(\mathbf{z}_t; t \in \mathbb{Z})$ est un b.b.s. vectoriel, la *série* $\sum_{t=1}^{\infty} \mathbf{z}_t$ *converge en moyenne* lorsque $n \rightarrow \infty$ ssi

$$\sum_{t=1}^{\infty} \|\mathbf{z}_t\|^2 < \infty$$

et alors

$$\left\| \sum_{t=1}^{\infty} \mathbf{z}_t \right\|^2 = \sum_{t=1}^{\infty} \|\mathbf{z}_t\|^2$$

(théorème de Pythagone).

Démonstration. Par exemple, Loève (1963, p.456). \square

Les Propositions 2.37 et 2.38 constituent le théorème de décomposition de Wold-Cramér (cas stationnaire : Wold (1938), cas général : Cramér (1961)) ou représentation canonique du processus du second ordre qui est la base des chapitres suivants.

Définition 2.34. Soient :

$$\begin{aligned} V_t &= V_2(\mathbf{z}; t), \text{ le sous-espace linéaire fermé sous-tendu par l'ensemble} \\ &\{(\mathbf{z}_t)_i; i = 1, \dots, k, s \leq t\}; \\ V_{-\infty} &= V_2(\mathbf{z}; -\infty) = \bigcap_{t \in \mathbb{Z}} V_2(\mathbf{z}; t); \\ V_{+\infty} &= V_2(\mathbf{z}) = V_2(\mathbf{z}; \infty) \end{aligned}$$

Proposition 2.35. $\forall t \in \mathbb{Z} : V_{t-1} \subseteq V_t$.

Démonstration. Les v.a. qui sous-tendent V_{t-1} appartiennent aussi à V_t , donc toute combinaison linéaire incluant ces variables appartient à V_t . Puisque V_t est fermé, toute limite d'une suite de Cauchy dans V_{t-1} appartient à V_t . \square

Définition 2.36. $(\mathbf{z}_t; t \in \mathbb{Z})$ est un *processus purement indéterminable* (p.p.i.) ssi $V_{-\infty} = \{\mathbf{0}\}$. $(\mathbf{z}_t; t \in \mathbb{Z})$ est un *processus déterminable* (p.d.) ssi $V_{-\infty} = V_{\infty}$.

Exemples.

(a) Si $(\mathbf{e}_t; t \in \mathbb{Z})$ est un b.b.n. vectoriel, alors un p.a.s. vectoriel $(\mathbf{z}_t; t \in \mathbb{Z})$ défini par

$$\mathbf{z}_t = \phi \mathbf{z}_{t-1} + \mathbf{e}_t$$

(ce qui implique comme on le verra plus loin que les valeurs propres de ϕ sont inférieures à 1 en module) est un p.p.i. En effet, pour tout $i = 1, \dots, k$, $(\mathbf{z}_s)_i \in V_2(\mathbf{e}; t)$, $\forall s \leq t$, d'où $V_2(\mathbf{z}; t) \subset V_2(\mathbf{e}; t)$. Inversement $(\mathbf{e}_s)_i \in V_2(\mathbf{z}; t)$, donc $V_2(\mathbf{z}; t) = V_2(\mathbf{e}; t)$. Or $\bigcap_{t \in \mathbb{Z}} V_2(\mathbf{e}; t) = \{\mathbf{0}\}$ ce qui termine la preuve.

(b) Si \mathbf{z}^* est une v.a.s. vectorielle, alors $(\mathbf{z}_t; t \in \mathbb{Z})$ défini par $\mathbf{z}_t = \mathbf{f}_t^\top \mathbf{z}^*$, où \mathbf{f}_t est une fonction arbitraire de \mathbb{Z} dans \mathbb{R}^k , est un p.d. En effet, $\forall t : V_2(\mathbf{z}; t)$ est le sous-espace de Hilbert sous-tendu par \mathbf{z}^* et donc $V_2(\mathbf{z}; -\infty) = V_2(\mathbf{z}; t) = V_2(\mathbf{z})$.

Proposition 2.37. *Tout p.a.s. vectoriel $(\mathbf{z}_t; t \in \mathbb{Z})$ est tel qu'il existe un p.p.i. unique $(\mathbf{z}'_t; t \in \mathbb{Z})$ et un p.d. unique $(\mathbf{z}''_t; t \in \mathbb{Z})$ tels que $\forall t \in \mathbb{Z} :$*

$$\mathbf{z}_t = \mathbf{z}'_t + \mathbf{z}''_t \text{ et } V_2(\mathbf{z}') \perp V_2(\mathbf{z}'').$$

Démonstration. C'est une application directe du théorème de décomposition (Proposition 2.18), composante par composante. \square

Proposition 2.38. (Décomposition de Wold-Cramér). *A tout p.p.i. vectoriel $(\mathbf{z}_t; t \in \mathbb{Z})$ correspond un p.a. $(\mathbf{e}_t; t \in \mathbb{Z})$ appelée processus innovation tel que*

1. $(\mathbf{e}_t; t \in \mathbb{Z})$ est un b.b.s. vectoriel, avec une matrice de covariance notée Σ_t ;
2. $\forall t \in \mathbb{Z} : \mathbf{e}_t \perp V_{t-1}$.

Notons F_t le sous-espace de Hilbert sous-tendu par les composantes $(\mathbf{e}_i)_i$ de \mathbf{e}_t ($i = 1, \dots, k$). Alors $V_t = V_{t-1} \oplus F_t$. De plus, pour tout t on a

$$\mathbf{z}_t = \mathbf{e}_t + \sum_{j=1}^{\infty} \Psi_{jt} \mathbf{e}_{t-j} \quad (2.2)$$

où les matrices Ψ_{jt} sont telles que la série (2.2) converge en moyenne, c'est-à-dire

$$\sum_{j=1}^{\infty} \Psi_{jt} \Sigma_{t-j} \Psi_{jt}^{\top} < \infty. \quad (2.3)$$

Démonstration. Soient $\hat{\mathbf{z}}_t = P(\mathbf{z}_t | V_{t-1})$ et $\mathbf{e}_t = \mathbf{z}_t - \hat{\mathbf{z}}_t$. Par construction, les conditions (a) et (b) sont vérifiées. Puisque $\mathbf{e}_{t-1} \perp V_{t-2}$, le théorème de décomposition 2.18 permet d'écrire composante par composante: $\hat{\mathbf{z}}_t = \mathbf{z}_t + \mathbf{z}''_t$ où $\mathbf{z}'_t, \mathbf{z}''_t \in V_{t-1}$, $\mathbf{z}'_t \in F_{t-1}$, $\mathbf{z}''_t \in V_{t-2}$. Par conséquent, il existe une matrice Ψ_{1t} telle que $\mathbf{z}'_t = \Psi_{1t} \mathbf{e}_{t-1}$ et donc

$$\mathbf{z}_t = \mathbf{e}_t + \Psi_{1t} \mathbf{e}_{t-1} + \mathbf{z}''_t.$$

Par récurrence, on a donc

$$\mathbf{z}_t = \mathbf{e}_t + \sum_{j=1}^{\infty} \Psi_{jt} \mathbf{e}_{t-j}.$$

Puisque $\mathbf{z}_t \in V_2(\mathbf{z})$, la série converge en moyenne. La Proposition 2.33 exprime ceci par la condition

$$\sum_{j=1}^{\infty} \|\Psi_{jt} \mathbf{e}_{t-j}\|^2 < \infty.$$

ce qui équivaut à (2.3). □

Remarques.

1. Puisque $\mathbf{e}_t = \mathbf{z}_t - P(\mathbf{z}_t | V_{t-1})$, l'innovation \mathbf{e}_t apparaît comme l'erreur de prévision d'horizon 1 faite à l'instant $t-1$ pour l'instant t et Σ_t est la matrice de covariance de cette erreur.
2. Si $(\mathbf{z}_t; t \in \mathbb{Z})$ est un p.p.i. vectoriel, les δ_{jt} ($j = 0, \dots, m$) où m est fini sont des matrices, $\delta_{0t} = I$, alors

$$\left(\sum_{j=0}^m \delta_{jt} \mathbf{z}_{t-j}; t \in \mathbb{Z} \right)$$

est un p.p.i. qui a même processus innovation que $(\mathbf{z}_t; t \in \mathbb{Z})$.

3. La décomposition (2.2) est unique mais les coefficients Ψ_{jt} ne sont uniquement déterminés que si les sous-espaces F_t sont de dimension k .

4. Si $(\mathbf{z}_t; t \in \mathbb{Z})$ est un p.a.s. admettant la décomposition donnée dans la Proposition 2.37, on définit son processus innovation comme étant le processus innovation de $(\mathbf{z}'_t; t \in \mathbb{Z})$.

Exemple. (De Schutter, (1977)) Supposons un p.p.i. bivarié $(\mathbf{z}_t; t \in \mathbb{Z})$, où $\mathbf{z}_t = (x_t, y_t)^\top$ est tel que $x_t = \beta_t y_t + e_t$ où β_t est une fonction du temps, $(e_t; t \in \mathbb{Z})$ est un b.b.s. non corrélé avec $(y_t; t \in \mathbb{Z})$ et de variance $\sigma_{e,t}^2$. Montrons comment on peut déterminer la décomposition de Wold-Cramér de $(\mathbf{z}_t; t \in \mathbb{Z})$. Remarquons d'abord que

$$V_{t-1} = V_2(y; t-1) \oplus V_2(e; t-1).$$

Notons $\mathbf{c}_t = (a_t, b_t)^\top$ l'innovation en t de \mathbf{z}_t . Il vient immédiatement que

$$\begin{aligned} b_t &= y_t - P(y_t | V_{t-1}) \\ &= y_t - P(y_t | V_2(y; t-1)) \end{aligned}$$

est l'innovation en t de $(y_t; t \in \mathbb{Z})$. D'autre part

$$\begin{aligned} a_t &= \beta_t y_t + e_t - P(\beta_t y_t + e_t | V_{t-1}) \\ &= \beta_t y_t + e_t - \beta_t P(y_t | V_2(y; t-1)) \\ &= \beta_t b_t + e_t. \end{aligned}$$

Par conséquent

$$\boldsymbol{\Sigma}_t = \begin{pmatrix} \beta_t^2 \sigma_{b,t}^2 + \sigma_{e,t}^2 & \beta_t \sigma_{b,t}^2 \\ \beta_t \sigma_{b,t}^2 & \sigma_{b,t}^2 \end{pmatrix},$$

où $\sigma_{b,t}^2$ est la variance de l'innovation de y_t . Si la décomposition de Wold-Cramér de $(y_t; t \in \mathbb{Z})$ est

$$y_t = b_t + \sum_{j=1}^{\infty} \psi_{jt} b_{t-j},$$

on obtient celle de $(\mathbf{z}_t; t \in \mathbb{Z})$ sous la forme

$$\mathbf{z}_t = \mathbf{c}_t + \sum_{j=1}^{\infty} \boldsymbol{\Psi}_{jt} \mathbf{c}_{t-j}$$

où

$$\boldsymbol{\Psi}_{jt} = \begin{pmatrix} 0 & \beta_t \psi_{jt} \\ 0 & \psi_{jt} \end{pmatrix}.$$

Chapitre 3

PROCESSUS ARMA ÉVOLUTIFS

Dans ce chapitre nous introduisons la classe des *processus* ARMA (processus autorégressif moyenne mobile = *autoregressive-moving average processes*), ainsi que les processus AR (*processus autorégressif*) et les processus MA (*processus moyenne mobile*). Notons que la terminologie traditionnelle est impropre puisqu'un processus ARMA général n'est ni autorégressif ni moyenne mobile. Pour éviter cette ambiguïté on utilisera le sigle anglais ARMA qui est d'usage universel.

Il faut noter que la définition de processus ARMA sous-entend généralement une idée de stationnarité qui sera seulement exploitée dans le chapitre 4. Ici on s'intéresse aux *processus* ARMA évolutifs encore appelés *processus* ARMA à coefficients dépendant du temps. Contrairement à des décompositions telles que celles présentées au chapitre 1, équation (1.3), qui est autorégressive d'ordre infini, ou celle du théorème de Wold-Cramér du chapitre 2, équation (2.2), qui est moyenne mobile d'ordre infini, on utilise une décomposition comportant un nombre fini de termes, donc un nombre fini de coefficients. C'est ce nombre fini de coefficients (fonctions du temps dans le chapitre 3, constantes dans le chapitre 4) qu'il s'agit de déterminer pour constituer ce qu'on appelle un *modèle* ARMA de la s.c. étudiée.

Les processus ARMA évolutifs ont été très peu étudiés dans la littérature, que ce soit du point de vue théorique ou du point de vue pratique. Citons néanmoins Quérouille (1957), Whittle (1965), Abdrabbo et Priestley (1967), Miller (1968, 1969), Rao (1970), Wegman (1974). Mélard (1975) a envisagé la condition nécessaire *d'indéterminabilité pure*, a généralisé la notion *d'inversibilité* et en a obtenu une condition nécessaire et suffisante. Les sections 3.2 et 3.4 contiennent une version améliorée de ces résultats. Le problème d'inversibilité a également été étudié par Hallin et Mélard (1977) et Hallin (1978, 1986). Hallin (1984 et les références citées dans cet article) a étudié le problème de la *factorisation de la fonction de covariance* : partant de la fonction de covariance d'un processus AR ou MA évolutif, il s'agit de déterminer les coefficients de celui-ci (en fait : les différents ensembles de coefficients qui donnent lieu à la même fonction de covariance). Le problème est ardu parce qu'on suppose que le processus démarre en $t = -\infty$, et non à un instant fini comme Anderson, Moore et Loo (1969). Notons toutefois qu'on connaît rarement la fonction de covariance d'un processus à partir de $t = -\infty$. De plus,

on est évidemment incapable d'estimer la fonction de covariance d'un processus ARMA évolutif à partir d'une s.c. unique, à moins que cette fonction de covariance ne dépende d'un nombre fini de paramètres. Hallin (1984) discute également les problèmes liés à la *multiplicité des représentations* ARMA dès le moment où les \mathbf{e}_t constituent un bruit blanc quelconque et pas nécessairement le processus innovation du processus étudié. Ici on supposera toujours que les \mathbf{e}_t sont les innovations du processus et c'est dans ce sens qu'on aura unicité de la représentation ARMA du processus. Voici quelques références récentes sur le sujet : Kozin et Nakajima (1980), Dunsmuir (1981), Tyssedal et Tjøstheim (1983), Bordignon et Masarotto (1983), Hamdi (1982).

3.1. Définitions et notations

Définition 3.1. Un p.p.i. vectoriel $(\mathbf{z}_t; t \in \mathbb{Z})$ avec $(\mathbf{e}_t; t \in \mathbb{Z})$ pour processus innovation est dit ARMA évolutif (ou ARMA à coefficients dépendant du temps) s'il vérifie l'équation

$$\mathbf{z}_t = \phi_{1t}\mathbf{z}_{t-1} + \dots + \phi_{pt}\mathbf{z}_{t-p} + \mathbf{e}_t - \theta_{1t}\mathbf{e}_{t-1} - \dots - \theta_{qt}\mathbf{e}_{t-q}, \quad (3.1)$$

où les coefficients matriciels ϕ_{it} ($i = 1, \dots, p$), θ_{it} ($i = 1, \dots, q$) et $\|\mathbf{e}_t\|^2 = \Sigma_t$ sont des fonctions du temps et p et q sont des constantes.

La partie autorégressive (AR) est $\phi_{1t}\mathbf{z}_{t-1} + \dots + \phi_{pt}\mathbf{z}_{t-p}$ tandis que la partie moyenne mobile (MA) est $-\theta_{1t}\mathbf{e}_{t-1} - \dots - \theta_{qt}\mathbf{e}_{t-q}$. Quand on veut insister sur les valeurs respectives de p et q , on qualifie le processus de ARMA(p, q) ou ARMA d'ordre (p, q). Un processus AR(p) est un processus ARMA($p, 0$) tandis qu'un processus MA(q) est un processus ARMA($0, q$). On imposera parfois que le processus ARMA est d'ordre minimal (p, q) c'est-à-dire $\phi_{pt} \neq \mathbf{0}$ et $\theta_{qt} \neq \mathbf{0}$ pour au moins au t . Notons \mathbf{I}_k la matrice unité d'ordre k .

Définition 3.2. Soit B l'opérateur de retard à droite, tel que $f_t B g_t = f_t g_{t-1}$. On peut alors réécrire l'équation (3.1) sous la forme $\phi_t(B)\mathbf{z}_t = \theta_t(B)\mathbf{e}_t$ où $\phi_t(B)$ est l'opérateur AR (à droite)

$$\phi_t(B) = \mathbf{I}_k - \phi_{1t}B + \dots + \phi_{pt}B^p$$

et $\theta_t(B)$ est l'opérateur MA (à droite)

$$\theta_t(B) = \mathbf{I}_k - \theta_{1t}B - \dots - \theta_{qt}B^q.$$

Des opérateurs tels que $\phi_t(B)$ et $\theta_t(B)$ sont appelés *opérateurs polynomiaux de retard à droite* (o.p.r.d.). Un o.p.r.d. est dit *constant* si $\phi_{t+h}(B) = \phi_t(B)$ pour tout h entier. La multiplication de deux o.p.r.d. est définie mais n'est pas commutative sauf si les deux o.p.r.d. sont constants et si $k = 1$.

Fréquemment, surtout dans le chapitre 4, on commettra l'abus de langage qui consiste à considérer B à la fois comme opérateur agissant dans V_2 et comme variable complexe, selon le contexte. Il n'en résultera aucune ambiguïté.

Définition 3.3. L'o.p.r.d. $\nabla = 1 - B$ est l'*opérateur de différence ordinaire* ("regular differencing operator"). L'o.p.r.d. $\nabla_s = 1 - B^s$, s entier positif, est l'*opérateur de différence saisonnière*. Plus généralement, on peut considérer un *opérateur cercle unité* $U(B)$ dont les zéros sont sur le cercle unité du plan complexe.

Définition 3.4. S'il existe deux entiers positifs d et D tels que $d + D > 0$ et tels que $(\nabla^d \nabla_s^D z_t; t \in \mathbb{Z})$ est un p.p.i. ARMA évolutif, alors $(z_t; t \in \mathbb{Z})$ est dit *processus ARIMA évolutif* ("integrated autoregressive-moving average process").

Cette définition donnée ici pour un processus scalaire peut s'étendre aux processus vectoriels en utilisant des vecteurs \mathbf{d} et \mathbf{D} . Les processus ARIMA sont un cas particulier des *processus ARUMA* utilisant un opérateur cercle unité (Anderson, 1979a).

3.2. Condition nécessaire pour un processus purement indéterminable

Pour la commodité des notations, on remplace ici p et q par l'infini en introduisant des coefficients $\phi_{it} = \mathbf{0}$ pour $i > p$ et $\theta_{it} = \mathbf{0}$ pour $i > q$, avec $\theta_{0t} = -\mathbf{I}_k$. Toutes les sommes considérées sont donc finies.

Proposition 3.5. Des conditions nécessaires pour qu'un processus ARMA évolutif vérifiant (3.1) soit un p.p.i. sont que

$$(a) \text{ l.i.m. } \sum_{j=m+1}^{\infty} (\alpha_{jt}^{(m)} \mathbf{z}_{t-j} + \psi_{jt}^{(m)} \mathbf{e}_{t-j}) = \mathbf{0},$$

$$(b) \text{ la série } \sum_{m=1}^{\infty} \psi_{mt}^{(m)} \Sigma_{t-m} \psi_{mt}^{(m)\top} \text{ converge,}$$

où les coefficients $\alpha_{jt}^{(m)}$ et $\psi_{jt}^{(m)}$ sont définis par récurrence :

$$\left. \begin{aligned} \psi_{jt}^{(0)} &= -\theta_{jt} & (j \geq 1), & \quad \psi_{0t}^{(0)} = \mathbf{I}_k, \\ \alpha_{jt}^{(0)} &= \phi_{jt} & (j \geq 1), \\ \psi_{jt}^{(m)} &= \psi_{jt}^{(m-1)} - \alpha_{mt}^{(m-1)} \theta_{j-m,t-m} & (j \geq m), \\ \alpha_{jt}^{(m)} &= \alpha_{jt}^{(m-1)} + \alpha_{mt}^{(m-1)} \phi_{j-m,t-m} & (j \geq m+1) \end{aligned} \right\} (m \geq 1).$$

Démonstration. L'équation (3.1) peut s'écrire

$$\mathbf{z}_t = \mathbf{z}'_t{}^{(0)} + \mathbf{z}''_t{}^{(0)},$$

où

$$\mathbf{z}'_t{}^{(0)} = \psi_{0t}^{(0)} \mathbf{e}_t,$$

$$\mathbf{z}_t^{(0)} = \sum_{j=1}^{\infty} \alpha_{jt}^{(0)} \mathbf{z}_{t-j} + \sum_{j=1}^{\infty} \psi_{jt}^{(0)} \mathbf{e}_{t-j}.$$

Montrons par récurrence que

$$\mathbf{z}_t = \mathbf{z}_t^{(i)} + \mathbf{z}_t^{(i)}, \quad (3.2)$$

où

$$\begin{aligned} \mathbf{z}_t^{(i)} &= \mathbf{e}_t + \sum_{j=1}^i \psi_{jt}^{(j)} \mathbf{e}_{t-j}, \\ \mathbf{z}_t^{(i)} &= \sum_{j=i+1}^{\infty} \alpha_{jt}^{(i)} \mathbf{z}_{t-j} + \sum_{j=i+1}^{\infty} \psi_{jt}^{(i)} \mathbf{e}_{t-j} \end{aligned}$$

pour tout $i \geq 0$. En effet, si (3.2) est vrai pour $i = m - 1$, on a

$$\mathbf{z}_t = \mathbf{z}_t^{(m-1)} + \alpha_{mt}^{(m-1)} \mathbf{z}_{t-m} + \psi_{mt}^{(m-1)} \mathbf{e}_{t-m} + \sum_{j=m+1}^{\infty} \alpha_{jt}^{(m-1)} \mathbf{z}_{t-j} + \sum_{j=m+1}^{\infty} \psi_{jt}^{(m-1)} \mathbf{e}_{t-j}.$$

En remplaçant \mathbf{z}_{t-m} au moyen de (3.1), il vient

$$\begin{aligned} \mathbf{z}_t &= \mathbf{z}_t^{(m-1)} + (\alpha_{mt}^{(m-1)} + \psi_{mt}^{(m-1)}) \mathbf{e}_{t-m} + \alpha_{mt}^{(m-1)} \left(\sum_{j=1}^{\infty} \phi_{j,t-m} \mathbf{z}_{t-j-m} - \sum_{j=1}^{\infty} \theta_{j,t-m} \mathbf{e}_{t-j-m} \right) \\ &\quad + \sum_{j=m+1}^{\infty} \alpha_{jt}^{(m-1)} \mathbf{z}_{t-j} + \sum_{j=m+1}^{\infty} \psi_{jt}^{(m-1)} \mathbf{e}_{t-j} \\ &= \mathbf{e}_t + \sum_{j=1}^{m-1} \psi_{jt}^{(j)} \mathbf{e}_{t-j} + \psi_{mt}^{(m)} \mathbf{e}_{t-m} + \sum_{j=m+1}^{\infty} \left(\alpha_{jt}^{(m-1)} + \alpha_{mt}^{(m-1)} \phi_{j-m,t-m} \right) \mathbf{z}_{t-j} \\ &\quad + \sum_{j=m+1}^{\infty} \left(\psi_{jt}^{(m-1)} - \alpha_{mt}^{(m-1)} \theta_{j-m,t-m} \right) \mathbf{e}_{t-j} \\ &= \mathbf{e}_t + \sum_{j=1}^m \psi_{jt}^{(j)} \mathbf{e}_{t-j} + \sum_{j=m+1}^{\infty} \alpha_{jt}^{(m)} \mathbf{z}_{t-j} + \sum_{j=m+1}^{\infty} \psi_{jt}^{(m)} \mathbf{e}_{t-j} = \mathbf{z}_t^{(m)} + \mathbf{z}_t^{(m)}. \end{aligned}$$

Puisque les \mathbf{e}_t sont les innovations des \mathbf{z}_t , les coefficients $\psi_{mt}^{(m)}$ sont donc les coefficients de la décomposition de Wold-Cramér du processus (Proposition 2.38). Donc l.i.m. $\mathbf{z}_t^{(m)} = \mathbf{z}_t$ pour $m \rightarrow \infty$, ce qui entraîne (b). Il s'ensuit que l.i.m. $\mathbf{z}_t^{(m)} = 0$ lorsque $m \rightarrow \infty$, donc (a). \square

Remarques.

1. Inversement, si des coefficients ϕ_{jt} et θ_{jt} , tels que (b) est vrai sont donnés, on peut définir par (3.1) un p.p.i. à partir d'un bruit blanc ($\mathbf{e}_t; t \in \mathbb{Z}$) qui en deviendra le processus innovation.

2. Les $\alpha_{mt}^{(m)}$ constituent les valeurs de la fonction de Green unilatérale associée à l'opérateur aux différences $\phi_t(B)$ (Hallin et Mélard, 1977 et Hallin, 1978).

Proposition 3.6. *La condition nécessaire pour qu'un processus AR(1), c'est-à-dire vérifiant la relation*

$$\mathbf{z}_t = \phi_t \mathbf{z}_{t-1} + \mathbf{e}_t$$

soit un p.p.i. est

$$\sum_{j=1}^{\infty} \left(\prod_{i=0}^{j-1} \phi_{t-i} \right) \boldsymbol{\Sigma}_{t-j} \left(\prod_{i=0}^{j-1} \phi_{t-i} \right)^{\top} < \infty.$$

Démonstration. Les relations de la Proposition 3.5 s'écrivent

$$\boldsymbol{\Psi}_{mt}^{(m)} = \prod_{i=0}^{m-1} \phi_{t-i} \quad \boldsymbol{\Psi}_{jt}^{(m)} = \mathbf{0} \quad (j > m).$$

Il en résulte que

$$\mathbf{z}'_t{}^{(m)} = \mathbf{e}_t + \sum_{j=1}^m \left(\prod_{i=0}^{j-1} \phi_{t-i} \right) \mathbf{e}_{t-j}.$$

On impose alors que l.i.m. $\mathbf{z}'_t{}^{(m)} = \mathbf{z}_t$, ce qui donne la condition énoncée. \square

Remarquons que dans le cas univarié, la condition s'écrit simplement

$$\sum_{j=1}^{\infty} \left(\prod_{i=0}^{j-1} \phi_{t-i} \right)^2 < \infty.$$

Proposition 3.7. *Pour un processus ARMA(1, p) évolutif, les coefficients définis dans la Proposition 3.5 sont donnés par les expressions suivantes*

$$\begin{aligned} \alpha_{m+1,t}^{(m)} &= \prod_{i=0}^m \phi_{1,t-j}, \\ \alpha_{j,t}^{(m)} &= \mathbf{0} \quad (j \geq m+2), \\ \boldsymbol{\Psi}_{jt}^{(m)} &= - \sum_{h=0}^m \left(\prod_{\ell=0}^{m-h-\ell} \phi_{1,t-\ell} \right) \boldsymbol{\theta}_{j-m+h,t-m+h} \quad (j = m, \dots, m+q), \\ \boldsymbol{\Psi}_{jt}^{(m)} &= \mathbf{0} \quad \text{pour} \quad j > m+q. \end{aligned}$$

Démonstration. Ces relations sont valables pour $m = 0$. Montrons que si elles sont valables pour m , elles le sont pour $m + 1$. En effet,

$$\alpha_{m+2,t}^{(m+1)} = \alpha_{m+2,t}^{(m)} + \alpha_{m+1,t}^{(m)} \phi_{1,t-m-1} = \left(\prod_{j=0}^m \phi_{1,t-j} \right) \phi_{1,t-m-1} = \prod_{j=0}^{m+1} \phi_{1,t-j},$$

$$\alpha_{j,t}^{(m+1)} = 0 \quad \text{pour } j \geq m+3.$$

Pour $j = m+1, \dots, m+q$:

$$\begin{aligned} \psi_{jt}^{(m+1)} &= \psi_{jt}^{(m)} - \alpha_{m+1,t}^{(m)} \theta_{j-m-1,t-m-1} \\ &= - \sum_{h=0}^m \left(\prod_{\ell=0}^{m-h-1} \phi_{1,t-\ell} \right) \theta_{j-m+h,t-m+h} - \left(\prod_{\ell=0}^m \phi_{1,t-j} \right) \theta_{j-m-1,t-m-1} \\ &= - \sum_{h=-1}^m \left(\prod_{\ell=0}^{m-h-1} \phi_{1,t-\ell} \right) \theta_{j-m+h,t-m+h} \\ &= - \sum_{h=0}^{m+1} \left(\prod_{\ell=0}^{m-h} \phi_{1,t-\ell} \right) \theta_{j-(m+1)+h,t-(m+1)+h} \end{aligned}$$

tandis que pour $j = m+q+1$,

$$\psi_{m+q+1,t}^{(m+1)} = - \left(\prod_{j=0}^m \phi_{1,t-j} \right) \theta_{q,t-(m+1)}$$

est donné par la même expression avec le seul terme correspondant à $h = 0$. \square

3.3. Caractérisation des processus AR et MA

Proposition 3.8. (Caractérisation des processus AR). *Un processus AR(p) vérifiant l'équation*

$$\mathbf{z}_t - \sum_{j=1}^p \phi_{jt} \mathbf{z}_{t-j} = \mathbf{e}_t$$

est tel que

$$\Gamma_{ts} - \sum_{j=1}^p \phi_{jt} \Gamma_{t-j,s} = \mathbf{0} \quad \text{pour } s < t. \quad (3.3)$$

Démonstration.

$$\begin{aligned} \Gamma_{ts} &= \langle \mathbf{z}_t, \mathbf{z}_s \rangle = \left\langle \sum_{j=1}^p \phi_{jt} \mathbf{z}_{t-j} + \mathbf{e}_t, \mathbf{z}_s \right\rangle \\ &= \sum_{j=1}^p \phi_{jt} \langle \mathbf{z}_{t-j}, \mathbf{z}_s \rangle + \langle \mathbf{e}_t, \mathbf{z}_s \rangle = \sum_{j=1}^p \phi_{jt} \Gamma_{t-j,s} + \langle \mathbf{e}_t, \mathbf{z}_s \rangle. \end{aligned}$$

Mais, pour $s < t$, $\mathbf{e}_t \perp V_2(\mathbf{z};s)$, donc $\langle \mathbf{e}_t, \mathbf{z}_s \rangle = 0$. \square

L'équation (3.3) est connue sous le nom d'équation de Yule-Walker. La généralisation considérée ici est due à de Falguerolles (1979) et Hallin et Ingenbleek (1983).

Définition 3.9. Un p.a.s. est dit *q-dépendant* si $\Gamma_{ts} = 0$ pour tous t, s tels que $|t-s| > q$.

Notons qu'un processus *q-dépendant* est *q'-dépendant* pour tout $q' > q$.

Proposition 3.10. *Un processus MA(q) est q-dépendant.*

Proposition 3.11. (Caractérisation des processus MA). *Un processus q-dépendant pour lequel Σ_t est inversible pour tout t est un processus MA(q).*

Démonstration. Par hypothèse, pour tout $s < t - q$, $\mathbf{z}_t \perp V_2(\mathbf{z}; s)$ d'où $\mathbf{z}_t \perp \mathbf{a}_s$. Considérons la décomposition de Wold-Cramér, Proposition 2.38. On en déduit que

$$\langle \mathbf{z}_t, \mathbf{e}_s \rangle = \sum_{j=1}^{\infty} \boldsymbol{\psi}_{jt} \langle \mathbf{e}_{t-j}, \mathbf{e}_s \rangle + \langle \mathbf{e}_t, \mathbf{e}_s \rangle$$

d'où, pour $s < t - q$:

$$\mathbf{0} = \boldsymbol{\psi}_{t-s,t} \Sigma_s.$$

Puisque Σ_s est inversible, il vient $\boldsymbol{\psi}_{jt} = \mathbf{0}$ pour $j > q$, ce qui montre que le processus est MA(q). \square

Pour une autre démonstration de cette proposition, n'utilisant pas la décomposition de Wold-Cramér, voir Hallin (1984). Le cas ARMA est étudié par Hamdi (1982).

3.4. Inversibilité

Définition 3.12. Un p.p.i. $(\mathbf{z}_t; t \in \mathbb{Z})$ est *inversible* ssi il existe des coefficients $\boldsymbol{\pi}_{jt}$ tels que les innovations \mathbf{e}_t s'obtiennent par

$$\mathbf{e}_t = \sum_{j=0}^{\infty} \boldsymbol{\pi}_{jt} \mathbf{z}_{t-j},$$

où la série converge en moyenne.

Remarquons que si on reprend la définition de processus ARMA sans imposer que les \mathbf{e}_t soient les innovations, et si néanmoins les \mathbf{e}_t peuvent s'obtenir comme série convergente en moyenne des \mathbf{z}_t présents et passés, alors les \mathbf{e}_t sont automatiquement les innovations. Cependant, ainsi que Wold (1938) l'a montré dans le cas des processus ARMA constants, les \mathbf{e}_t peuvent être les innovations sans que la condition d'inversibilité soit vérifiée. En effet, tout élément de V_t ne peut pas nécessairement s'exprimer comme série en les \mathbf{z}_{t-j} . Tout ce qu'on peut affirmer, c'est qu'on peut obtenir tout élément de V_t , en particulier \mathbf{e}_t , comme limite d'une suite de Cauchy de v.a. qui sont les sommes de séries en les \mathbf{z}_{t-j} qui convergent en moyenne. Voir aussi Hallin (1986).

Proposition 3.13. *La limite en moyenne pour $m \rightarrow \infty$ de*

$$\sum_{j=0}^m \boldsymbol{\pi}_{jt}^{(j)} \mathbf{z}_{t-j}$$

existe et est finie ssi

$$\text{l.i.m.}_{m \rightarrow \infty} \sum_{j=m+1}^{\infty} (\boldsymbol{\pi}_{jt}^{(m)} \mathbf{z}_{t-j}) + \boldsymbol{\beta}_{jt}^{(m)} \mathbf{e}_{t-j} = \mathbf{0}$$

où les coefficients $\boldsymbol{\pi}_{jt}^{(m)}$ et $\boldsymbol{\beta}_{jt}^{(m)}$ sont définis par récurrence

$$\left. \begin{aligned} \boldsymbol{\pi}_{jt}^{(0)} &= -\boldsymbol{\phi}_{jt} & (j \geq 1), & \quad \boldsymbol{\pi}_{0t}^{(0)} = \mathbf{I}_k, & \quad \boldsymbol{\beta}_{jt}^{(0)} = \boldsymbol{\theta}_{jt} & \quad (j \geq 1) \\ \boldsymbol{\pi}_{jt}^{(m)} &= \boldsymbol{\pi}_{jt}^{(m-1)} - \boldsymbol{\beta}_{mt}^{(m-1)} \boldsymbol{\phi}_{j-m,t-m} & (j \geq m) \\ \boldsymbol{\beta}_{jt}^{(m)} &= \boldsymbol{\beta}_{jt}^{(m-1)} + \boldsymbol{\beta}_{mt}^{(m-1)} \boldsymbol{\theta}_{j-m,t-m} & (j \geq m+1) \end{aligned} \right\} (m \geq 1).$$

où l'on note $\boldsymbol{\phi}_{0t} = -\mathbf{I}_k$. L'une ou l'autre des conditions entraîne que le p.p.i. est inversible.

Démonstration. La démonstration est similaire à celle de la Proposition 3.5. On montre que

$$\mathbf{e}_t = \mathbf{e}_t^{\prime(i)} + \mathbf{e}_t^{\prime\prime(i)}$$

pour tout i , où

$$\mathbf{e}_t^{\prime(i)} = \sum_{j=0}^i \boldsymbol{\pi}_{jt}^{(j)} \mathbf{z}_{t-j},$$

$$\mathbf{e}_t^{\prime\prime(i)} = \sum_{j=i+1}^{\infty} \boldsymbol{\pi}_{jt}^{(i)} \mathbf{z}_{t-j} + \sum_{j=i+1}^{\infty} \boldsymbol{\beta}_{jt}^{(i)} \mathbf{e}_{t-j}.$$

C'est vrai pour $i = 0$. Si ce l'est pour $i = m - 1$, montrons que ce l'est encore pour $i = m$. En effet,

$$\begin{aligned} \mathbf{e}_t &= \mathbf{e}_t^{\prime(m-1)} + (\boldsymbol{\pi}_{mt}^{(m-1)} + \boldsymbol{\beta}_{mt}^{(m-1)}) \mathbf{z}_{t-m} + \sum_{j=m+1}^{\infty} \left(\boldsymbol{\pi}_{jt}^{(m-1)} - \boldsymbol{\beta}_{mt}^{(m-1)} \boldsymbol{\phi}_{j-m,t-m} \right) \mathbf{z}_{t-j} \\ &\quad + \sum_{j=m+1}^{\infty} \left(\boldsymbol{\beta}_{jt}^{(m-1)} + \boldsymbol{\beta}_{mt}^{(m-1)} \boldsymbol{\theta}_{j-m,t-m} \right) \mathbf{e}_{t-j} \\ &= \mathbf{e}_t^{\prime(m)} + \mathbf{e}_t^{\prime\prime(m)}. \end{aligned} \tag{3.4}$$

Puisque \mathbf{e}_t existe, il est équivalent que l.i.m. $\mathbf{e}_t^{\prime(m)} = \mathbf{e}_t$ ou l.i.m. $\mathbf{e}_t^{\prime\prime(m)} = \mathbf{0}$, lorsque $m \rightarrow \infty$. Cette dernière condition est donc une condition suffisante d'inversibilité, compte tenu de la forme prise par $\mathbf{e}_t^{\prime(m)}$. \square

Proposition 3.14. *Une condition suffisante pour qu'un p.p.i. MA(1), défini par l'équation*

$$\mathbf{z}_t = \mathbf{e}_t - \boldsymbol{\theta}_t \mathbf{e}_{t-1}$$

où les \mathbf{e}_t sont les innovations, soit inversible est

$$\lim_{m \rightarrow \infty} \left(\prod_{j=0}^{m-1} \boldsymbol{\theta}_{t-j} \right) \boldsymbol{\Sigma}_{t-m} \left(\prod_{j=0}^{m-1} \boldsymbol{\theta}_{t-j} \right)^\top = \mathbf{0}.$$

Démonstration. Il est aisé de voir que la décomposition (3.4) prend la forme

$$\mathbf{e}_t = \mathbf{z}_t + \boldsymbol{\theta}_t \mathbf{z}_{t-1} + \boldsymbol{\theta}_t \boldsymbol{\theta}_{t-1} \mathbf{z}_{t-2} + \dots + \left(\prod_{j=0}^{m-1} \boldsymbol{\theta}_{t-j} \right) \mathbf{z}_{t-m} - \left(\prod_{j=0}^m \boldsymbol{\theta}_{t-j} \right) \mathbf{e}_{t-m-1}.$$

D'où

$$\mathbf{e}_t''^{(m)} = - \left(\prod_{j=0}^m \boldsymbol{\theta}_{t-j} \right) \mathbf{e}_{t-m-1}.$$

En exprimant que l.i.m. $\mathbf{e}_t''^{(m)} = \mathbf{0}$ pour $m \rightarrow \infty$, on obtient la condition suffisante ci-dessus. \square

3.5. Exemples de processus ARMA évolutifs

Une sous-classe intéressante de processus ARMA évolutifs est constituée des processus FARMAG.

Définition 3.15. Un p.p.i. $(\mathbf{z}_t; t \in \mathbb{Z})$ est dit FARMAG s'il vérifie l'équation

$$\boldsymbol{\phi}(B) \mathbf{F}_t^{-1} \mathbf{z}_t = \boldsymbol{\theta}(B) \mathbf{G}_t \mathbf{b}_t \quad (3.5)$$

où

$$\begin{aligned} \boldsymbol{\phi}(B) &= \mathbf{I}_k - \boldsymbol{\phi}_1 B - \dots - \boldsymbol{\phi}_p B^p, \\ \boldsymbol{\theta}(B) &= \mathbf{I}_k - \boldsymbol{\theta}_1 B - \dots - \boldsymbol{\theta}_q B^q, \end{aligned}$$

\mathbf{F}_t et \mathbf{G}_t sont des matrices inversibles pour tout t ; le processus $(\mathbf{b}_t; t \in \mathbb{Z})$ est un b.b.s.c. de vecteur moyenne $\mathbf{0}$ et de matrice de covariance $\boldsymbol{\Sigma}$, tel que $\mathbf{b}_t \perp V_{t-1}$. Si $\mathbf{F}_t = \mathbf{I}_k$ pour tout t , le processus est dit ARMAG. Si $\mathbf{G}_t = \mathbf{I}_k$ pour tout t , le processus est dit FARMA. Plus généralement, si $(\nabla^d \nabla_s^D \mathbf{z}_t; t \in \mathbb{Z})$ est un processus FARMAG, on dit que $(\mathbf{z}_t; t \in \mathbb{Z})$ est un processus FARIMAG.

Le processus est évidemment un processus ARMA évolutif représenté par (3.1), où l'on pose

$$\boldsymbol{\phi}_t(B) = \boldsymbol{\phi}(B) \mathbf{F}_t^{-1} \quad , \quad \boldsymbol{\theta}_t(B) = \boldsymbol{\theta}(B) \mathbf{G}_t.$$

Puisque $\mathbf{b}_t \perp V_{t-1}$, l'innovation en t , \mathbf{e}_t , vaut $\mathbf{F}_t \mathbf{G}_t \mathbf{b}_t$ et la matrice de covariance de cette innovation est $\mathbf{F}_t \mathbf{G}_t \boldsymbol{\Sigma} \mathbf{G}_t^\top \mathbf{F}_t^\top$.

Les processus FARMAG ont été considérés incidemment par Herbst (1963) et étudiés par Mélard (1975, 1977).

Dans ce qui suit, on considère l'inverse d'o.p.r.d. constant $\phi(B)$ et $\theta(B)$, qu'on note respectivement $\phi^{-1}(B)$ et $\theta^{-1}(B)$. Ce sont les matrices inverses de matrices de polynômes en B , définies de façon formelle. Ce sont donc des séries formelles en B .

Proposition 3.16. *La décomposition de Wold-Cramér d'un processus FARMAG peut s'écrire comme suit :*

$$\mathbf{z}_t = \mathbf{F}_t \phi^{-1}(B) \theta(B) \mathbf{G}_t \mathbf{b}_t. \quad (3.6)$$

Si la condition d'inversibilité est satisfaite, on peut écrire similairement

$$\mathbf{b}_t = \mathbf{G}_t^{-1} \theta^{-1}(B) \phi(B) \mathbf{F}_t^{-1} \mathbf{z}_t. \quad (3.7)$$

Démonstration. (a) Considérons $\phi^{-1}(B)$ et appliquons cet o.p.r.d. aux deux membres de (3.5). Par définition, $\phi^{-1}(B) \phi(B) = \mathbf{I}_k$, d'où

$$\mathbf{F}_t^{-1} \mathbf{z}_t = \phi^{-1}(B) \theta(B) \mathbf{G}_t \mathbf{b}_t.$$

Le second membre est une série convergence en moyenne puisque $\mathbf{F}_t^{-1} \mathbf{z}_t$ est un p.p.i. Il suffit de prémultiplier les deux membres de cette équation par \mathbf{F}_t pour trouver (3.6).

(b) On procède de même avec l'o.p.r.d. $\theta^{-1}(B)$ et on prémultiplie par \mathbf{G}_t^{-1} pour trouver (3.7). \square

Notons que les manipulations d'opérateurs effectuées pour arriver à (3.6) et (3.7) ne sont pas valables dans le cas général de (3.1).

Les exemples ci-dessous concernent d'autres approches de l'analyse des s.c. univariées ($k = 1$) où les modèles sous-jacents sont des processus ARMA évolutifs. On se limite au cas univarié. Les résultats 3.19, 3.21, 3.23 et 3.25 ont été établis par McKenzie (1974, 1976) avec toutefois la réserve mentionnée dans la Remarque 1 qui suit la Proposition 3.19. Les démonstrations sont toutefois différentes.

Définition 3.17. Considérons le *modèle de prévision* défini à la section 1.5, avec $\ell = 1$ pour horizon,

$$z_t = \hat{z}_{t-1}(1) + e_t$$

où l'on note $\hat{z}_{t-1}(1)$ la prévision d'horizon 1 (désignée à la section 1.5 par $\tilde{\mathcal{F}}_t$). On utilise la seule hypothèse H_4 . Le modèle de prévision est dit *sous-entendu par un processus ARIMA évolutif* si les observations z_t vérifient la relation

$$\phi_t(B) \nabla^d \nabla_s^D z_t = \theta_t(B) e_t.$$

Définition 3.18. Le *modèle de lissage exponentiel d'ordre k* est défini (Brown, 1959 et Goodman, 1974) par

$$\hat{z}_{t-1}(1) = \sum_{i=1}^k (-1)^{i-1} \binom{k}{i} z_t^{[i]}$$

où les $z_t^{[i]}$ sont les valeurs lissées d'ordre i , définies par les relations de récurrence

$$z_t^{[i]} = \alpha z_{t-1}^{[i-1]} + (1 - \alpha) z_{t-1}^{[i]} \quad (3.8)$$

et $z_t^{[0]} = z_t$.

Proposition 3.19. *Le modèle de lissage exponentiel d'ordre k est sous-tendu par le processus ARIMA évolutif*

$$\nabla^k z_t = [1 - (1 - \alpha)B]^k e_t,$$

si la condition d'inversibilité est remplie.

Démonstration. On peut écrire (3.8) sous la forme

$$[1 - (1 - \alpha)B]z_t^{[i]} = \alpha B z_t^{[i-1]}.$$

Introduisons l'o.p.r.d. $\alpha(B) = 1 - (1 - \alpha)B$. Il est simple de montrer que

$$\alpha^i(B) z_t^{[i]} = \alpha^i B^i z_t.$$

Par définition, $z_t = \hat{z}_{t-1}(1) + e_t$, d'où

$$\begin{aligned} \alpha^k(B) z_t &= \sum_{i=1}^k (-1)^{i-1} \binom{k}{i} \alpha^k(B) z_t^{[i]} + \alpha^k(B) e_t \\ &= \left[\sum_{i=1}^k (-1)^{i-1} \binom{k}{i} \alpha^{k-i}(B) \alpha^i B^i \right] z_t + \alpha^k(B) e_t, \end{aligned}$$

ou encore

$$\left[\sum_{i=0}^k (-1)^i \binom{k}{i} \alpha^{k-i}(B) \alpha^i B^i \right] z_t = \alpha^k(B) e_t.$$

L'o.p.r.d. du membre de gauche peut s'écrire

$$[\alpha(B) - \alpha B]^k = (1 - B)^k = \nabla^k$$

et la proposition en résulte immédiatement. \square

Remarques.

1. Le caractère évolutif du processus est fort limité. En effet, l'hypothèse H_3 du chapitre 1 n'étant pas imposée, e_t est de variance σ_t^2 quelconque, pas nécessairement constante. On a donc en fait un processus ARIMAG. Il en sera de même pour les cas qui suivent.

2. La méthode de prévision correspondant à ce modèle est utilisée avec des conditions initiales $z_1^{[i]} = z_1$ pour tout i , par exemple. Ces conditions initiales sont parfois discutables. Nous préférons les ignorer et écrire les relations pour tout $t \in \mathbb{Z}$. La même remarque vaut pour les modèles qui suivent.

3. Si σ_t^2 est constant, la condition d'inversibilité est simplement $0 < \alpha < 2$. Autrement, par application de la Proposition 3.14, cette condition est $\lim_{m \rightarrow \infty} (1 - \alpha)^{2m} \sigma_{t-m}^2 = 0$.

Définition 3.20. Le modèle de lissage exponentiel avec tendance (par exemple, Holt et al., 1963) est défini par les relations:

$$\hat{z}_{t-1}(1) = n_t + p_t,$$

$$n_t = \alpha z_{t-1} + (1 - \alpha)(n_{t-1} + p_{t-1}), \quad (3.9)$$

$$p_t = \beta(n_t - n_{t-1}) + (1 - \beta)p_{t-1}. \quad (3.10)$$

Proposition 3.21. Le modèle de lissage exponentiel avec tendance est sous-tendu par le processus ARIMA évolutif

$$\nabla^2 z_t = [1 - (2 - \alpha - \alpha\beta)B + (1 - \alpha)B^2]e_t$$

si la condition d'inversibilité est remplie.

Démonstration. De la relation $z_t = n_t + p_t + e_t$, nous tirons

$$\nabla^2 z_t = \nabla^2 n_t + \nabla^2 p_t + \nabla^2 e_t.$$

Or (3.9) et (3.10) impliquent d'abord

$$\nabla n_t = \alpha(z_{t-1} - n_{t-1} - p_{t-1}) + p_{t-1} = \alpha e_{t-1} + p_{t-1},$$

$$\nabla p_t = \beta \nabla n_t - \beta p_{t-1} = \alpha \beta e_{t-1},$$

puis

$$\nabla^2 n_t = \alpha \nabla B e_t + \nabla B p_t = (\alpha \nabla B + \alpha \beta B^2) e_t,$$

$$\nabla^2 p_t = \alpha \beta \nabla B e_t.$$

En rassemblant les résultats, on obtient

$$\nabla^2 z_t = (\alpha \nabla B + \alpha \beta B^2 + \alpha \beta \nabla B + \nabla^2) e_t,$$

ce qui peut s'écrire comme dans l'énoncé de la proposition. \square

Définition 3.22. Le modèle de lissage exponentiel de Winters (1960) est défini par les relations

$$\hat{z}_{t-1}(1) = n_t + p_t + s_{t+1-r},$$

$$n_t = \alpha(z_{t-1} - s_{t-r}) + (1 - \alpha)(n_{t-1} + p_{t-1}), \quad (3.11)$$

$$p_t = \beta(n_t - n_{t-1}) + (1 - \beta)p_{t-1} \quad (3.12)$$

$$s_t = \gamma(z_{t-1} - n_t) + (1 - \gamma)s_{t-r}, \quad (3.13)$$

où r est un entier supérieur à 2.

Proposition 3.23. *Le modèle de lissage exponentiel de Winters est sous-tendu par le processus ARIMA évolutif*

$$\nabla \nabla_r z_t = [1 - (1 - \alpha - \alpha\beta)B + \alpha\beta \sum_{i=1}^{r-1} B^i - (1 - \alpha\beta + \alpha\gamma - \gamma)B^r + (1 - \alpha)(1 - \gamma)B^{r+1}]e_t$$

pourvu que la condition d'inversibilité soit vérifiée.

Démonstration. De la relation $z_t = n_t + p_t + s_{t+1-r} + e_t$, nous tirons

$$\nabla \nabla_r z_t = \nabla \nabla_r n_t + \nabla \nabla_r p_t + \nabla \nabla_r B^{r-1} s_t + \nabla \nabla_r e_t. \quad (3.14)$$

Or (3.11) à (3.13) impliquent que

$$\nabla n_t = \alpha e_{t-1} + p_{t-1},$$

$$\nabla p_t = \beta(\nabla n_t - p_{t-1}) = \alpha\beta B e_t,$$

$$\nabla_r s_t = \gamma(z_{t-1} - n_t - s_{t-r}) = \gamma(e_{t-1} + p_{t-1} - \nabla n_t) = \gamma(1 - \alpha)B e_t.$$

Ensuite,

$$\nabla \nabla_r p_t = \alpha\beta \nabla_r B e_t,$$

$$\nabla \nabla_r s_t = \gamma(1 - \alpha) \nabla B e_t,$$

$$\nabla \nabla_r n_t = \alpha \nabla_r B e_t + \nabla_r B p_t = \alpha \nabla_r B e_t + \left(B \sum_{i=0}^{r-1} B^i \right) \nabla p_t = [\alpha \nabla_r B + \alpha\beta B^2 \sum_{i=0}^{r-1} B^i] e_t.$$

En insérant ces expressions dans (3.14), on trouve l'équation annoncée. \square

Définition 3.24. *Le modèle de prévision par moyenne mobile simple d'ordre m est défini par*

$$\hat{z}_{t-1}(1) = \frac{1}{m} \sum_{i=1}^m z_{t-i}.$$

Proposition 3.25. *Le modèle de prévision par moyenne mobile simple d'ordre m est sous-tendu par le modèle ARIMA évolutif*

$$\left(1 + \sum_{i=1}^m \frac{m-i}{m} B^i \right) \nabla z_t = e_t.$$

Démonstration. Par définition, on a

$$z_t = \hat{z}_{t-1}(1) + e_t = \frac{1}{m} \sum_{i=1}^m B^i z_t + e_t.$$

Le modèle sous-tendu s'écrit donc

$$\left(1 - \frac{1}{m} \sum_{i=1}^m B^i \right) z_t = e_t$$

ou

$$\left(1 + \sum_{i=1}^{m-1} \frac{m-i}{m} B^i\right) \nabla z_t = e_t$$

comme annoncé. \square

Définition 3.26. Le modèle à deux aléas général est défini par les relations

$$z_t = n_t + x_t,$$

$$n_t = n_{t-1} + \mu_t + y_t,$$

$$x_t = \rho_1 x_{t-1} + b_t,$$

$$y_t = \rho_2 y_{t-1} + c_t,$$

où $b_t, c_t \perp b_s, c_s$ pour $t \neq s$ et $E(b_t) = E(c_t) = 0, \forall t$ et μ_t est une fonction du temps.

Proposition 3.27. Le modèle à deux aléas général est sous-tendu par un processus ARMA(2,2) évolutif relatif à $\nabla z_t - \mu_t$:

$$[1 - (\rho_1 + \rho_2)B + \rho_1 \rho_2 B^2](\nabla z_t - \mu_t) = e_t - \theta_{1t} e_{t-1} - \theta_{2t} e_{t-2}.$$

Démonstration. En effet:

$$\nabla z_t - \mu_t = y_t + \nabla x_t,$$

$$(1 - \rho_1 B)(\nabla z_t - \mu_t) = (1 - \rho_1 B)y_t + \nabla b_t,$$

$$(1 - \rho_2 B)(1 - \rho_1 B)(\nabla z_t - \mu_t) = (1 - \rho_1 B)c_t + (1 - \rho_2 B)\nabla b_t.$$

Il est évident que $(1 - \rho_1 B)c_t \perp (1 - \rho_2 B)\nabla b_s$ dès que $|t - s| > 2$. Le second membre représente donc un processus 2-dépendant dont la fonction de covariance dépend de la covariance entre b_t et c_t et des variances de b_t et c_t . La proposition résulte alors de la Proposition 3.11. \square

Remarque. Le modèle original de Bachelet et Morlat (1966) suppose que $b_t \perp c_t$ et que b_t et c_t sont de variances constantes. Dans ce cas, θ_{1t} et θ_{2t} sont constants. Les formules exprimant ces coefficients sont données par Libert (1977). Si de plus $\rho_1 = \rho_2 = 0$, on retrouve le modèle du lissage exponentiel d'ordre 1. Si $\rho_1 = 0, \rho_2 = 1$ et $\mu_t = 0$, on retrouve le modèle de Theil et Wage étudié par Van Winkel (1976).

Définition 3.28. Le modèle linéaire dynamique (m.l.d.) vectoriel est défini par les relations suivantes, pour $t \geq 1$

$$\mathbf{z}_t = \mathbf{H}_t \mathbf{w}_t + \mathbf{a}_t, \quad (3.15)$$

$$\mathbf{w}_t = \mathbf{F}_t \mathbf{w}_{t-1} + \mathbf{G}_t \mathbf{e}_t, \quad (3.16)$$

où $\mathbf{z}_t, \mathbf{w}_t, \mathbf{a}_t$ et \mathbf{e}_t sont des v.a. vectorielles de dimensions respectives k, r, k et m , $\mathbf{H}_t, \mathbf{F}_t$ et \mathbf{G}_t sont des matrices non aléatoires de dimensions respectives $k \times r, r \times r$

et $r \times m$, $(\mathbf{e}_t; t \geq 1)$ et $(\mathbf{a}_t; t \geq 1)$ sont des b.b.s. vectoriels de moyennes nulles et tels que $\|\mathbf{e}_t\|^2 = \mathbf{\Sigma}_t$, $\|\mathbf{a}_t\|^2 = \mathbf{S}_t$, et $\mathbf{e}_t \perp \mathbf{a}_s$, $\forall t, s$, \mathbf{w}_0 est de moyenne nulle et $\|\mathbf{w}_0\|^2 = \mathbf{\Sigma}_0$. \mathbf{w}_t est appelé le *vecteur d'état*, (3.15) est appelée l'*équation d'observation* et (3.16) est appelée l'*équation de transition*.

Le m.l.d., aussi appelé système stochastique dynamique linéaire, a été étudié principalement par les ingénieurs, à des fins de filtrage, c'est-à-dire de l'estimation du vecteur d'état \mathbf{w}_t à partir des observations présentes et passées \mathbf{z}_{t-i} , $i \geq 0$. Les premiers développements sont dus à Kalman (1960). Harrison et Stevens (1976) ont utilisé le m.l.d. dans le cadre d'une méthode de prévision bayésienne.

Proposition 3.29. *Le processus décrit dans la Définition 3.28 est un p.p.i.*

Démonstration. Cela résulte directement de la Définition 2.36. □

Proposition 3.30. *Si \mathbf{H}_t est une fonction périodique en t , de période s , et si $\mathbf{F}_t = \mathbf{I}_r$, alors $(\nabla_s \mathbf{z}_t; t \geq 1)$ est un processus s -dépendant.*

Démonstration. En appliquant l'opérateur de différence saisonnière à \mathbf{z}_t , il vient

$$\nabla_s \mathbf{z}_t = \mathbf{H}_t \nabla_s \mathbf{w}_t + \nabla_s \mathbf{a}_t.$$

Mais (3.16) entraîne, compte tenu de l'hypothèse sur \mathbf{F}_t , que

$$\mathbf{w}_t = \mathbf{w}_{t-s} + \sum_{j=0}^{s-1} \mathbf{G}_{t-j} \mathbf{e}_{t-j}$$

d'où

$$\nabla_s \mathbf{z}_t = \mathbf{H}_t \sum_{j=0}^{s-1} \mathbf{G}_{t-j} \mathbf{e}_{t-j} + \mathbf{a}_t - \mathbf{a}_{t-s}. \quad (3.17)$$

Les p.a. $(\mathbf{x}'_t; t > s)$ et $(\mathbf{x}''_t; t > s)$ définis par

$$\mathbf{x}'_t = \sum_{j=0}^{s-1} \mathbf{G}_{t-j} \mathbf{e}_{t-j}, \quad \mathbf{x}''_t = \mathbf{a}_t - \mathbf{a}_{t-s}$$

sont respectivement MA($s-1$) et MA(s) et non corrélés entre eux. Le second membre de (3.17) définit donc un processus s -dépendant. □

Remarque. Le modèle vectoriel à saisonnier additif et le modèle à croissance constante et saisonnier multiplicatif (Harrison et Stevens, 1976) sont tous deux des cas particuliers de la classe envisagée ici.

Proposition 3.31. *Si $m = r$ et si les matrices \mathbf{H}_t , \mathbf{F}_t et \mathbf{G}_t satisfont aux conditions suivantes*

$$\mathbf{H}_t = \begin{bmatrix} 1 & 0 & \dots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 1 & 0 & \dots & 0 \end{bmatrix}, \quad \mathbf{F}_t = \mathbf{G}_t = \begin{bmatrix} 1 & 1 & \dots & 1 & 1 \\ 0 & 1 & & & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \dots & \dots & 0 & 1 \end{bmatrix}$$

alors $(\nabla^m \mathbf{z}_t; t > m)$ est un processus m -dépendant.

Démonstration. Pour tout j tel que $1 \leq j \leq m$, on a

$$(\mathbf{w}_t)_j = \sum_{i=j}^m (\mathbf{w}_{t-1})_i + \sum_{i=j}^m (\mathbf{e}_t)_i,$$

d'où

$$(\mathbf{w}_t)_j - (\mathbf{w}_{t-1})_j = \sum_{i=j+1}^m (\mathbf{w}_{t-1})_i + \sum_{i=j}^m (\mathbf{e}_t)_i$$

ou

$$\nabla (\mathbf{w}_t)_j = (\mathbf{w}_t)_{j+1} + (\mathbf{e}_t)_j$$

pour $1 \leq j \leq m-1$ tandis que $\nabla (\mathbf{w}_t)_m = (\mathbf{a}_t)_m$. Par ailleurs,

$$\nabla^m \mathbf{z}_t = \nabla^m (\mathbf{w}_t)_1 \mathbf{1} + \nabla^m \mathbf{a}_t \text{ avec } \mathbf{1}^\top = (1, \dots, 1).$$

On définit $\mathbf{x}'_t = \nabla^m (\mathbf{w}_t)_1$, ce qui peut s'écrire

$$\mathbf{x}'_t = \sum_{i=0}^{m-1} \nabla^i (\mathbf{e}_t)_{m-i}$$

d'où $(\mathbf{x}'_t; t > m)$ est m -dépendant. Il en est de même de $(\mathbf{x}''_t; t > m)$ avec $\mathbf{x}''_t = \nabla^m \mathbf{a}_t$. Finalement

$$\nabla^m \mathbf{z}_t = \left[\sum_{i=0}^{m-1} \nabla^i (\mathbf{e}_t)_{m-i} \right] \mathbf{1} + \nabla^m \mathbf{a}_t$$

est également m -dépendant. \square

Les modèles envisagés ici sont les modèles de croissance polynomiale (Harrison, 1967).

Proposition 3.32. *Un processus ARMA(p, q) évolutif peut se représenter comme solution du m.l.d. vectoriel.*

Démonstration. Cela peut se faire de plusieurs manières, par exemple la suivante. Soit $r = kR$ où $R = \max(p, q + 1)$. Soit $\mathbf{H}_t = (\mathbf{I}_k, \mathbf{0}_k, \dots, \mathbf{0}_k)$ et $\mathbf{a}_t = \mathbf{0}$, où $\mathbf{0}_k$ est une matrice $k \times k$ de zéros. On a donc $\mathbf{z}_t = (\mathbf{w}_t)_1$ comme équation d'observation. Dés lors

$$(\mathbf{w}_t)_1 = \sum_{j=1}^p \phi_{jt} \mathbf{z}_{t-j} + \mathbf{e}_t - \sum_{j=1}^q \theta_{jt} \mathbf{e}_{t-j} = \phi_{1t} (\mathbf{w}_{t-1})_1 + (\mathbf{w}_{t-1})_2 + \mathbf{e}_t,$$

où

$$(\mathbf{w}_t)_2 = \sum_{j=2}^p \phi_{j,t+1} \mathbf{z}_{t+1-j} - \sum_{j=1}^q \theta_{j,t+1} \mathbf{e}_{t+1-j}.$$

De même

$$(\mathbf{w}_t)_2 = \phi_{2,t+1} (\mathbf{w}_{t-1})_1 - \theta_{1,t+1} \mathbf{e}_t + (\mathbf{w}_{t-1})_3,$$

où

$$(\mathbf{w}_t)_3 = \sum_{j=3}^p \phi_{j,t+2} \mathbf{z}_{t+2-j} - \sum_{j=2}^q \theta_{j,t+2} \mathbf{e}_{t+2-j},$$

et ainsi de suite jusqu'à $(\mathbf{w}_t)_R$. Si l'on note $\phi_{jt} = \mathbf{0}$ pour $p < j \leq R$ et $\theta_{jt} = \mathbf{0}$ pour $q < j \leq R-1$, on peut écrire $\mathbf{w}_t = \mathbf{F}_t \mathbf{W}_{t-1} + \mathbf{G}_t \mathbf{e}_t$ où

$$\mathbf{F}_t = \begin{bmatrix} \phi_{1t} & \vdots & & & & \\ \phi_{2,t+1} & \vdots & \mathbf{I}_{(R-1)k} & & & \\ \dots & \vdots & & & & \\ \dots & \vdots & & & & \\ \phi_{R,t+R-1} & \vdots & & & \mathbf{0} & \end{bmatrix}, \quad \mathbf{G}_t = \begin{bmatrix} \mathbf{I}_k & & & & & \\ -\theta_{1,t+1} & & & & & \\ \vdots & & & & & \\ -\theta_{R-1,t+R-1} & & & & & \end{bmatrix}.$$

□

3.6. La fonction de covariance

Cette section concerne le calcul de la fonction de covariance Γ_{ts} d'un processus ARMA(p,q) évolutif. De la décomposition de Wold-Cramér

$$\mathbf{z}_t = \sum_{j=0}^{\infty} \psi_{jt} \mathbf{e}_{t-j}$$

où $\psi_{0t} = \mathbf{I}_k$, on déduit que

$$\begin{aligned} \Gamma_{ts} &= \sum_{j=0}^{\infty} \sum_{h=0}^{\infty} \psi_{jt} \langle \mathbf{e}_{t-j}, \mathbf{e}_{s-h} \rangle \psi_{hs}^{\top} \\ &= \sum_{j=t-s}^{\infty} \psi_{jt} \Sigma_{t-j} \psi_{s-t+j,s}^{\top}. \end{aligned} \quad (3.18)$$

Dans (3.18) et la suite, on suppose que $s \leq t$, ce qui n'est pas restrictif car $\Gamma_{st} = \Gamma_{ts}^{\top}$. Il est clair que (3.18) n'est pas pratique pour le calcul numérique. Si l'on suppose que la fonction de covariance est connue pour $s, t \leq 0$, mettons, on peut la déterminer pour tout s et t comme suit. Notons Γ'_{ts} la fonction de covariance du processus MA(q) ($\mathbf{y}_t; t \in \mathbb{Z}$) où

$$\mathbf{y}_t = \mathbf{e}_t - \sum_{j=1}^q \boldsymbol{\theta}_{jt} \mathbf{e}_{t-j}. \quad (3.19)$$

Elle est calculable par la formule

$$\boldsymbol{\Gamma}'_{ts} = \sum_{j=t-s}^q \boldsymbol{\theta}_{jt} \boldsymbol{\Sigma}_{t-j} \boldsymbol{\theta}_{s-t+j,s}^\top,$$

où $\boldsymbol{\theta}_{0t} = -\mathbf{I}_k$. On remarque que $\boldsymbol{\Gamma}'_{ts} = \mathbf{0}$ pour $t > s + q$. Introduisons $\boldsymbol{\Gamma}''_{ts} = \langle \mathbf{y}_t, \mathbf{z}_s \rangle$. Pour $t > s + q$, (3.19) ne fait intervenir que des variables orthogonales à \mathbf{z}_s , donc $\boldsymbol{\Gamma}''_{ts} = 0$. Sachant que

$$\mathbf{z}_s = \sum_{i=1}^p \boldsymbol{\phi}_{is} \mathbf{z}_{s-i} + \mathbf{y}_s,$$

il vient la relation de récurrence

$$\boldsymbol{\Gamma}''_{ts} = \sum_{i=1}^{\min(p, q-t+s)} \boldsymbol{\Gamma}''_{t, s-i} \boldsymbol{\phi}_{is}^\top + \boldsymbol{\Gamma}'_{ts}$$

qui permet de calculer successivement $\boldsymbol{\Gamma}''_{t, t+q}$, $\boldsymbol{\Gamma}''_{t, t+q-1}$, ..., $\boldsymbol{\Gamma}''_{tt}$. Enfin,

$$\begin{aligned} \boldsymbol{\Gamma}_{ts} &= \langle \mathbf{z}_t, \mathbf{z}_s \rangle \\ &= \langle \mathbf{y}_t + \sum_{i=1}^p \boldsymbol{\phi}_{it} \mathbf{z}_{t-i}, \mathbf{z}_s \rangle \\ &= \boldsymbol{\Gamma}''_{ts} + \sum_{i=1}^p \boldsymbol{\phi}_{it} \boldsymbol{\Gamma}_{t-i, s}. \end{aligned} \quad (3.20)$$

Puisque $\boldsymbol{\Gamma}_{ts}$ est supposé connu pour $t, s \leq 0$, cette relation permet de déterminer successivement les $\boldsymbol{\Gamma}_{1s}$, puis les $\boldsymbol{\Gamma}_{2s}$, ..., pour les valeurs requises de s , $s \geq 1$.

Insistons bien sur le fait que la connaissance de $\boldsymbol{\Gamma}_{ts}$ pour $s, t \leq 0$ est une condition suffisante mais pas nécessaire pour déterminer les valeurs suivantes (par exemple, Hallin et Ingenbleek, 1982, pour les processus AR). Cette section généralise au cas multivarié la méthode présentée par Mélard et Kiehm (1981).

3.7. La prévision

A la section 1.5 on avait développé une formulation du modèle de prévision qui envisageait simultanément plusieurs horizons. Ce modèle a été particularisé sous la forme du modèle multivarié qui ne fait plus intervenir que les prévisions d'horizon 1, par l'intermédiaire des innovations. On montre ici que le modèle multivarié est plus général qu'on pouvait le penser puisqu'il permet la détermination des prévisions à des horizons supérieurs à 1.

Notons $\hat{\mathbf{z}}_t(h)$ la prévision d'horizon h faite en t , $h \geq 1$, c'est-à-dire $P(\mathbf{z}_{t+h}|V_t)$.
Puisque

$$\mathbf{z}_{t+h} = \sum_{i=1}^p \boldsymbol{\phi}_{i,t+h} \mathbf{z}_{t+h-i} + \mathbf{e}_{t+h} - \sum_{j=1}^q \boldsymbol{\theta}_{j,t+h} \mathbf{e}_{t+h-j},$$

on peut écrire (Whittle, 1965)

$$\hat{\mathbf{z}}_t(h) = \sum_{i=1}^{\min(p,h-1)} \boldsymbol{\phi}_{i,t+h} \hat{\mathbf{z}}_t(h-i) + \sum_{i=h}^p \boldsymbol{\phi}_{i,t+h} \mathbf{z}_{t+h-i} - \sum_{j=h}^q \boldsymbol{\theta}_{j,t+h} \mathbf{e}_{t+h-j}.$$

Les prévisions $\hat{\mathbf{z}}_t(1)$, $\hat{\mathbf{z}}_t(2)$, ... sont donc calculables par récurrence.

La matrice de covariance des erreurs de prévisions d'horizon h est aisée à déterminer si on se réfère à la décomposition de Wold-Cramér du processus en $t+h$

$$\mathbf{z}_{t+h} = \mathbf{e}_{t+h} + \sum_{j=1}^{\infty} \boldsymbol{\psi}_{j,t+h} \mathbf{e}_{t+h-j}.$$

En effet, $\hat{\mathbf{z}}_t(h)$ s'écrit directement sous la forme

$$\hat{\mathbf{z}}_t(h) = \sum_{j=h}^{\infty} \boldsymbol{\psi}_{j,t+h} \mathbf{e}_{t+h-j},$$

d'où, par différence,

$$\mathbf{z}_{t+h} - \hat{\mathbf{z}}_t(h) = \mathbf{e}_{t+h} + \sum_{j=1}^{h-1} \boldsymbol{\psi}_{j,t+h} \mathbf{e}_{t+h-j}.$$

Par conséquent,

$$\|\mathbf{z}_{t+h} - \hat{\mathbf{z}}_t(h)\|^2 = \boldsymbol{\Sigma}_{t+h} + \sum_{j=1}^{h-1} \boldsymbol{\psi}_{j,t+h} \boldsymbol{\Sigma}_{t+h-j} \boldsymbol{\psi}_{j,t+h}^\top.$$

Chapitre 4

PROCESSUS ARMA CONSTANTS

Ce chapitre est consacré aux processus ARMA à coefficients constants, principalement aux processus univariés. Les résultats sont bien connus (par exemple, Quérouille, 1957, Hannan, 1970, Fuller, 1976, Rozanov, 1967 et Priestley, 1981). Pour la plupart, ils seront établis à partir du chapitre 3.

4.1. Définitions

Définition 4.1. Un p.a.s. est un processus stationnaire du second ordre (p.s.s.) ssi Γ_{ts} est fonction de $t - s$ seulement. On écrit $\gamma_h = \langle \mathbf{z}_t, \mathbf{z}_{t-h} \rangle = \Gamma_{t,t-h}$, qu'on appelle autocovariance d'ordre ou de retard h .

On appelle $(\gamma_h; h \in \mathbb{Z})$ la fonction d'autocovariance.

Proposition 4.2. *Le processus innovation d'un p.s.s. est un b.b.s.c. dont la matrice de covariance est notée Σ .*

Proposition 4.3. *La décomposition de Wold(-Cramér) d'un p.s.s. purement indéterminable (p.s.s.p.i) s'écrit*

$$\mathbf{z}_t = \sum_{j=0}^{\infty} \boldsymbol{\psi}_j \mathbf{e}_{t-j},$$

où les coefficients, $\boldsymbol{\psi}_j$ sont constants et $\boldsymbol{\psi}_0 = \mathbf{I}_k$. De plus,

$$\gamma_h = \sum_{j=h}^{\infty} \boldsymbol{\psi}_j \boldsymbol{\Sigma} \boldsymbol{\psi}_{j-h}^T = \boldsymbol{\gamma}_{-h}^T, \quad h \geq 0.$$

Définition 4.4. Un p.s.s.p.i. $(\mathbf{z}_t; t \in \mathbb{Z})$ avec $(\mathbf{e}_t; t \in \mathbb{Z})$ pour processus innovation est dit ARMA constant (ou ARMA à coefficients constants ou ARMA) s'il vérifie l'équation

$$\mathbf{z}_t = \phi_1 \mathbf{z}_{t-1} + \dots + \phi_p \mathbf{z}_{t-p} + \mathbf{e}_t - \theta_1 \mathbf{e}_{t-1} - \dots - \theta_q \mathbf{e}_{t-q} \quad (4.1)$$

avec $\|\mathbf{e}_t\| = \Sigma$, où ϕ_i ($i = 1, \dots, p$), θ_j ($j = 1, \dots, q$), Σ , p et q sont des constantes. On définit l'opérateur autorégressif (AR)

$$\phi(B) = \mathbf{I}_k - \phi_1 B - \dots - \phi_p B^p$$

et l'opérateur moyenne mobile

$$\theta(B) = \mathbf{I}_k - \theta_1 B - \dots - \theta_q B^q.$$

4.2. Stationnarité et inversibilité

Proposition 4.5. Si $(\mathbf{z}_t; t \in \mathbb{Z})$ est un p.s.s., le p.a. $(\mathbf{w}_t; t \in \mathbb{Z})$ défini par

$$\mathbf{w}_t = \sum_{j=0}^m \delta_j \mathbf{z}_{t-j}$$

est un p.s.s. quels que soient les coefficients δ_j constants ($j = 0, \dots, m$), où m est fini. De plus, si $(\mathbf{z}_t; t \in \mathbb{Z})$ est un p.s.s.p.i. et $\delta_0 = \mathbf{I}_k$, les processus innovations sont identiques.

Proposition 4.6. Une condition suffisante pour qu'il existe un p.s.s. qui soit solution de (4.1) est que les zéros du polynôme $\det[\phi(B)]$ soient supérieurs à 1 en module.

La démonstration utilise la Proposition 3.7 et les lemmes suivants.

Lemme 4.6.1 Un processus ARMA(p, q) est tel que (4.1) peut se mettre sous la forme ARMA(p, q)

$$\mathbf{Z}_t = \Phi \mathbf{Z}_{t-1} + \mathbf{E}_t - \Theta_1 \mathbf{E}_{t-1} - \dots - \Theta_q \mathbf{E}_{t-q}.$$

Démonstration. En effet, on peut poser $\mathbf{Z}_t^\top = (\mathbf{z}_t^\top, \mathbf{z}_{t-1}^\top, \dots, \mathbf{z}_{t-p+1}^\top)$, $\mathbf{E}_t^\top = (\mathbf{e}_t^\top, 0^\top, \dots, 0^\top)$ et les matrices $(kp) \times (kp)$

$$\Phi = \begin{bmatrix} \phi_1 & \phi_2 & \dots & \dots & \phi_{p-1} & \phi_p \\ \mathbf{I}_k & \mathbf{0} & \dots & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_k & \ddots & & \vdots & \vdots \\ & & \ddots & \ddots & \vdots & \vdots \\ & & & & \mathbf{0} & \vdots \\ \mathbf{0} & \dots & \dots & \mathbf{0} & \mathbf{I}_k & \mathbf{0} \end{bmatrix}, \quad \Theta_i = \begin{bmatrix} \theta_i & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \dots \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \dots & \mathbf{0} \end{bmatrix}. \quad (4.2)$$

□

Notons que la matrice de covariance de \mathbf{E}_t n'est pas inversible.

Lemme 4.6.2 Les valeurs propres de Φ sont les kp solutions de l'équation

$$\det[\mathbf{I}_k \lambda^p - \phi_1 \lambda^{p-1} - \dots - \phi_{p-1} \lambda - \phi_p] = 0.$$

Démonstration. En effet, l'équation caractéristique peut s'écrire

$$\det[\lambda \mathbf{I}_{kp} - \Phi] = 0 \quad \text{ou} \quad \det[\mathbf{J} + \mathbf{e} \phi^\top] = 0$$

avec

$$\mathbf{J} = \begin{bmatrix} \lambda \mathbf{I}_k & & & & \\ -\mathbf{I}_k & \lambda \mathbf{I}_k & & & 0 \\ & \ddots & \ddots & & \\ & & 0 & \ddots & \ddots \\ & & & & \mathbf{I}_k & \lambda \mathbf{I}_k \end{bmatrix}, \quad \mathbf{e} = \begin{bmatrix} \mathbf{I}_k \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}, \quad \phi = \begin{bmatrix} -\phi_1 \\ -\phi_2 \\ \vdots \\ -\phi_p \end{bmatrix}.$$

On peut évaluer un tel déterminant comme suit:

$$\det[\mathbf{J} + \mathbf{e} \phi^\top] = \det[\mathbf{J}] \det[\mathbf{I}_k + \phi^\top \mathbf{J}^{-1} \mathbf{e}].$$

Or, \mathbf{J} est une matrice triangulaire inférieure, donc

$$\det[\mathbf{J}] = \lambda^{kp}, \quad \mathbf{J}^{-1} = \begin{bmatrix} \lambda^{-1} \mathbf{I}_k & & & & \\ \lambda^{-2} \mathbf{I}_k & \ddots & & & 0 \\ & \vdots & \ddots & & \\ & & \lambda^{-p} \mathbf{I}_k & \dots & \lambda^{-2} \mathbf{I}_k & \lambda^{-1} \mathbf{I}_k \end{bmatrix}.$$

Par conséquent,

$$\mathbf{J}^{-1} \mathbf{e} = \begin{bmatrix} \lambda^{-1} \mathbf{I}_k \\ \lambda^{-2} \mathbf{I}_k \\ \vdots \\ \lambda^{-p} \mathbf{I}_k \end{bmatrix}$$

et

$$\det[\mathbf{I}_k + \phi^\top \mathbf{J}^{-1} \mathbf{e}] = \det[\mathbf{I}_k - \sum_{i=1}^p \phi_i \lambda^{-i}].$$

L'équation caractéristique s'écrit donc finalement

$$\det[\mathbf{I}_k \lambda^p - \sum_{i=1}^p \phi_i \lambda^{p-i}] = 0.$$

□

Lemme 4.6.3 Une condition suffisante pour que

$$\sum_{m=1}^{\infty} \boldsymbol{\phi}^m \boldsymbol{\Sigma} (\boldsymbol{\phi}^\top)^m < \infty$$

est que toutes les valeurs propres de $\boldsymbol{\phi}$ soient inférieures à 1 en module.

Démonstration. Soit $\varepsilon > 0$. Montrons qu'il existe un entier $M(\varepsilon)$ tel que pour tout $m_1, m_2 > M(\varepsilon)$ et pour tout $i, j = 1, \dots, k$, l'élément (i, j) de

$$\mathbf{P}(m_1, m_2) = \sum_{m=m_1}^{m_2} \boldsymbol{\phi}^m \boldsymbol{\Sigma} (\boldsymbol{\phi}^\top)^m$$

soit inférieur à ε . On peut écrire $\boldsymbol{\phi} = \mathbf{S}\mathbf{J}\mathbf{S}^{-1}$ où \mathbf{J} est une matrice de Jordan. Cette matrice a tous ses éléments (i, j) nuls sauf pour $j = i$ et $j = i + 1$. Les éléments diagonaux sont les valeurs propres λ_i de $\boldsymbol{\phi}$, rangées par ordre décroissant du module. De plus, l'élément $(i, i + 1)$ vaut 0 ou 1 selon que $\lambda_i \neq \lambda_{i+1}$ ou $\lambda_i = \lambda_{i+1}$. Dès lors

$$\boldsymbol{\phi}^m = (\mathbf{S}\mathbf{J}\mathbf{S}^{-1})^m = \mathbf{S}\mathbf{J}^m\mathbf{S}^{-1}.$$

On cherche à majorer les éléments de

$$\mathbf{P}(m_1, m_2) = \sum_{m=m_1}^{m_2} \mathbf{S}\mathbf{J}^m\mathbf{S}^{-1}\boldsymbol{\Sigma}(\mathbf{S}^{-1})^\top(\mathbf{J}^\top)^m\mathbf{S}^\top.$$

Or, les éléments de \mathbf{J}^m peuvent être majorés par une expression du type $\binom{m}{r}\lambda^{m-r}$, où $0 \leq r \leq k$ et $\lambda = \max[|\lambda_1|, \dots, |\lambda_k|]$. Puisque $\lambda < 1$ par hypothèse, on a, pourvu que $m > 2k$:

$$\binom{m}{r}\lambda^{m-r} \leq \binom{m}{k}\lambda^{m-k} \leq \frac{m^k \lambda^{m-k}}{k!}.$$

Posons $\mathbf{T} = \mathbf{S}^{-1}\boldsymbol{\Sigma}(\mathbf{S}^{-1})^\top$. Si \mathbf{X} est une matrice $k \times k$, d'éléments x_{ij} , notons $N(\mathbf{X}) = k \max_{i,j} |x_{ij}|$. Un majorant des éléments de $\mathbf{P}(m_1, m_2)$ est donc le suivant

$$\sum_{m=m_1}^{m_2} [N(\mathbf{S}) \frac{m^k}{k!} \lambda^{m-k}]^2 N(\mathbf{T}) \leq K \sum_{m=m_1}^{m_2} m^{2k} \lambda^{2m}$$

où K dépend de $\boldsymbol{\phi}$ et $\boldsymbol{\Sigma}$ mais pas de m_1, m_2 . Le rapport de deux termes successifs de cette somme vaut $[(m+1)/m]^{2k} \lambda^2$ qui tend vers λ^2 lorsque $m \rightarrow \infty$. Puisque $\lambda^2 < 1$, la série $\sum_m m^{2k} \lambda^{2m}$ converge, par application du critère de d'Alembert. On peut donc déterminer $M(\varepsilon)$ tel que pour tout $m_1, m_2 > M(\varepsilon)$ on ait

$$\sum_{m=m_1}^{m_2} m^{2k} \lambda^{2m} < \frac{\varepsilon}{K}$$

assurant ainsi que les éléments de $\mathbf{P}(m_1, m_2)$ sont majorés par ε . \square

Remarque. Si $\boldsymbol{\phi}$ est une matrice diagonalisable, c'est-à-dire si ses valeurs propres sont distinctes, la démonstration est plus rapide puisque \mathbf{J}^m est alors une matrice

diagonale constituée des λ_i^m . Les éléments de $\mathbf{P}(m_1, m_2)$ sont donc majorés par

$$\sum_{m=m_1}^{m_2} N(\mathbf{S})\lambda^m N(\mathbf{T})\lambda^m N(\mathbf{S}) \leq K_1 \sum_{m=m_1}^{m_2} \lambda^{2m}.$$

Le lemme résulte alors de la convergence de la série géométrique de raison λ^2 , lorsque $\lambda < 1$.

Démonstration. (de la Proposition 4.6). En vertu du Lemme 4.6.1, appliquons la Proposition 3.7 à un processus ARMA(1, q) constant en omettant les indices t inutiles. Il vient :

$$\begin{aligned} \alpha_{m+1}^{(m)} &= \Phi^{m+1}, & \alpha_j^{(m)} &= \mathbf{0} & (j \geq m+2), \\ \psi_j^{(m)} &= - \sum_{h=0}^m \Phi^{m-h} \Theta_{j-m+h}, & & & (j = m, \dots, m+q), \\ \psi_j^{(m)} &= \mathbf{0} & & & (j \geq m+q+1). \end{aligned}$$

Pour $m \geq q+1$ et $j = m, \dots, m+q$, on a l'expression plus simple

$$\psi_j^{(m)} = - \sum_{h=0}^{m+q-j} \Phi^{m-h} \Theta_{j-m+h}$$

d'où

$$\begin{aligned} \psi_m^{(m)} &= - \sum_{h=0}^q \Phi^{m-h} \Theta_h = \Phi^m - \sum_{h=1}^q \Phi^{m-h} \Theta_h \\ &= \Phi^{m-q} \left(\Phi^q - \sum_{h=1}^q \Phi^{q-h} \Theta_h \right). \end{aligned}$$

Plus généralement,

$$\psi_m^{(m)} = - \sum_{h=0}^{\min[m,q]} \Phi^{m-h} \Theta_h. \quad (4.3)$$

(Notons qu'on peut retrouver ce résultat en considérant les coefficients de la série formelle $(1 - \Phi B)^{-1} \Theta(B)$.) Considérons le p.a.

$$\mathbf{z}_t = \sum_{m=0}^{\infty} \psi_m^{(m)} \mathbf{e}_{t-m}. \quad (4.4)$$

Vérifions d'abord que la série converge en moyenne. On étudie à cette fin la série

$$\sum_{m=q+1}^{\infty} \psi_{mt}^{(m)} \Sigma \psi_{mt}^{(m)} = \sum_{m=q+1}^{\infty} \Phi^{m-q} \left(\Phi^q - \sum_{h=1}^q \Phi^{q-h} \Theta_h \right) \Sigma \left(\Phi^q - \sum_{h=1}^q \Phi^{q-h} \Theta_h \right)^{\top} (\Phi^{\top})^{m-q}. \quad (4.5)$$

Cette série est du type envisagé au Lemme 4.6.3. Elle converge donc si les valeurs propres de Φ sont inférieures à 1 en module. Or, le Lemme 4.6.2 montre que ces valeurs propres sont les inverses des solutions de l'équation déterminantale

$$\det[\mathbf{I}_k - \phi_1 B - \dots - \phi_p B^p] = 0.$$

Si ces solutions sont supérieures à 1 en module, la série (4.5) converge. Le processus défini par (4.4) est donc du second ordre. De plus, il est stationnaire et, par construction même, vérifie (4.1). \square

Remarques. La condition énoncée dans le théorème est suffisante mais pas nécessaire. Les deux raisons sont simplement illustrées :

(i) Si l'équation (4.1) peut être simplifiée parce que $\phi_p(B)$ et $\theta_q(B)$ contiennent des facteurs communs à gauche, la série converge sans que la condition ne doive être imposée sur les zéros correspondants. Par exemple, dans le cas ARMA(1, q), on peut avoir

$$\lambda^q \mathbf{I}_k - \sum_{h=1}^q \lambda^{q-h} \boldsymbol{\theta}_h = (\lambda \mathbf{I}_k - \phi_1) \left(\lambda^{q-1} \mathbf{I}_k - \sum_{h=1}^{q-1} \lambda^{q-1-h} \boldsymbol{\theta}_h \right),$$

de sorte qu'en remplaçant λ par ϕ_1 , il vient

$$\phi_1^q - \sum_{h=1}^q \phi_1^{q-h} \boldsymbol{\theta}_h = 0.$$

Cette absence de facteurs communs à gauche fait partie des conditions d'identifiabilité d'un modèle ARMA (Hannan, 1969).

(ii) L'exemple suivant montre une autre cause de problème, liée au rang de la matrice de covariance des innovations:

$$\mathbf{z}_t = \begin{bmatrix} 0.5 & 0 \\ 0 & 2 \end{bmatrix} \mathbf{z}_{t-1} + \mathbf{e}_t, \text{ avec } \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

En effet $(\mathbf{z}_t)_2 = 0$ est solution stationnaire de cette équation alors qu'une valeur propre de ϕ_1 vaut 2.

Dans le cas scalaire, il est aisé d'énoncer une condition nécessaire et suffisante.

Proposition 4.7. *Une condition nécessaire et suffisante pour qu'il existe une solution de (4.1) avec $k = 1$ qui soit un p.p.s.p.i en supposant qu'aucun zéro de $\phi_p(B)$ n'est zéro de $\theta_q(B)$ et que $\phi_p \neq 0$, est que les zéros du polynôme $\phi_p(B)$ soient supérieurs à 1 en module.*

Démonstration. Il reste à établir la condition nécessaire. On utilise directement la Proposition 3.5. Supposons que $\phi_p(B)$ soit factorisé en

$$\phi_p(B) = \prod_{\ell=1}^p (1 - \phi_\ell B)$$

où les φ_ℓ sont les inverses des zéros de $\phi_p(B)$ et sont réels ou constituent des paires de complexes imaginaires conjugués. De plus, les φ_ℓ sont différents de zéro. Choisissons arbitrairement $\ell \in \{1, \dots, p\}$ et posons

$$\phi^{(\ell)} = \prod_{\substack{j=1 \\ j \neq \ell}}^p (1 - \varphi_j B) = 1 - \sum_{j=1}^{p-1} \phi_j^{(\ell)} B^j.$$

On introduit le p.a. $\{w_t^{(\ell)}; t \in \mathbb{Z}\}$, à coefficients complexes, défini par

$$w_t^{(\ell)} = \phi^{(\ell)}(B)z_t.$$

C'est un p.p.s.p.i. en vertu de la Proposition 4.5. De plus, il vérifie l'équation

$$w_t^{(\ell)} = \varphi_\ell w_{t-1}^{(\ell)} + \theta(B)e_t.$$

Considérons les relations de la Proposition 3.7 déjà adaptées à un processus ARMA(1, q) constant dans la démonstration de la Proposition 4.7. En insérant les résultats dans la condition (b) de la Proposition 3.5, on exprime que la série

$$\sum_{m=0}^{\infty} \left(\sum_{h=0}^{\min[m, q]} \varphi_\ell^{m-h} \theta_h \right) e_{t-m}$$

converge en moyenne. Il est équivalent d'imposer la convergence de la série

$$\sum_{m=0}^{\infty} \varphi_\ell^m \left(\sum_{h=0}^q \varphi^{-h} \theta_h \right) e_{t-m} = \left(\sum_{h=0}^q \varphi^{-h} \theta_h \right) \sum_{m=0}^{\infty} \varphi_\ell^m e_{t-m}.$$

Mais φ_ℓ^{-1} n'est pas un zéro de $\theta_q(B)$, d'où

$$\sum_{h=0}^q \varphi_\ell^{-h} \theta_h \neq 0.$$

Il faut donc que $\sum \varphi_\ell^m e_{t-m}$ converge en moyenne, d'où $\sum |\varphi_\ell|^{2m} < \infty$, ce qui entraîne $|\varphi_\ell| < 1$. \square

Remarques.

(1) Nous ne connaissons pas, dans le cas vectoriel, de démonstration de la condition nécessaire qui ne fasse pas appel à l'approche spectrale et la théorie des matrices analytiques (Rozanov, 1967).

(2) La condition suffisante exprimée dans la Proposition 4.7 permet une démonstration alternative de la Proposition 4.6. En effet, si $\det[\phi_p(B)] \neq 0$, on peut introduire la matrice inverse $\phi_p^{-1}(B) = \phi_p^*(B) / \det[\phi_p(B)]$ et $\phi_p^*(B)$ est la matrice transposée des cofacteurs de $\phi_p(B)$. Appliquons $\phi_p^*(B)$ comme o.p.r.d. aux deux membres de (4.1). Il vient

$$\det[\phi_p(B)]z_t = \phi_p^*(B)\theta_q(B)e_t.$$

On applique alors la condition suffisante de la Proposition 4.7 composante par composante. Si les zéros de $\det[\phi_p(B)]$ sont supérieurs à 1 en module, on peut écrire la décomposition de Wold

$$z_t = \psi(B)e_t \quad \text{où} \quad \psi(B) = \phi_p^{-1}(B)\theta_q(B).$$

(3) La théorie des fonctions analytiques permet plus que de montrer que la convergence en moyenne c'est-à-dire $\sum(\psi_m^{(m)})^2 < \infty$. En effet, si $\phi_p(B)$ n'a aucun zéro pour $|B| \leq 1$, $\phi_p^{-1}(B)$ est une fonction analytique dans le cercle unité, donc également $\psi(B)$. La série entière converge pour $|B| \leq 1$, donc $\sum|\psi_m^{(m)}| < \infty$.

Proposition 4.8. *Si $z_t = \sum_{j=0}^{\infty} \psi_j e_{t-j}$ est la décomposition moyenne mobile infinie d'un p.s.s.p.i. scalaire, alors $\sum_{j=0}^{\infty} \psi_j B^j \neq 0$ pour tout B complexe tel que $|B| < 1$.*

Démonstration. Supposons au contraire qu'il existe b , $|b| < 1$, tel que $\sum_{j=0}^{\infty} \psi_j b^j = 0$. Puisque $\psi_0 = 1$, on a $b \neq 0$. Il s'ensuit que l'on peut écrire

$$\sum_{j=0}^{\infty} \psi_j B^j = (B-b) \sum_{j=0}^{\infty} \psi'_j B^j$$

avec $\psi'_0 = -1/b$. Définissons formellement

$$\sum_{j=0}^{\infty} \psi'_j B^j = (1-\bar{b}B) \sum_{j=0}^{\infty} \psi''_j B^j = \frac{1-\bar{b}B}{B-b} \sum_{j=0}^{\infty} \psi_j B^j.$$

Pour tout B de module 1, soit $B = e^{i\omega}$, on a $|B-b| = |1-b/B| = |1-b\bar{B}| = |1-\bar{b}B|$ d'où

$$\left| \sum_{j=0}^{\infty} \psi'_j e^{i\omega j} \right| = \left| \sum_{j=0}^{\infty} \psi_j e^{i\omega j} \right|.$$

La deuxième série $\psi(\omega)$ est une fonction de carré sommable sur $[-\pi, \pi[$ par rapport à la mesure de Lebesgue, puisque $\sum_{j=0}^{\infty} |\psi_j|^2 < \infty$. Il en est donc de même de la première série $\psi'(\omega)$ et donc $\sum_{j=0}^{\infty} |\psi'_j|^2 < \infty$. Considérons le p.a. $(z'_t; t \in \mathbb{Z})$ défini par $z'_t = \sum_{j=0}^{\infty} \psi'_j e_{t-j}$. C'est un p.s.s. et sa fonction d'autocovariance $(\gamma'_h; h \in \mathbb{Z})$ est obtenue par

$$\gamma'_h = \sigma^2 \sum_{j=h}^{\infty} \psi'_j \bar{\psi}'_{j-h} \quad \text{pour } h > 0.$$

Remarquons que $\bar{\psi}'(\omega)e^{-i\omega h} = \sum_{j=h}^{\infty} \psi'_j e^{-i\omega j}$. Par application de l'identité de Parseval, il vient

$$\begin{aligned}
\sum_{j=h}^{\infty} \psi'_j \bar{\psi}'_{j-h} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \psi'(\omega) \bar{\psi}'(\omega) e^{-i\omega h} d\omega \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} |\psi'(\omega)|^2 e^{-i\omega h} d\omega \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} |\psi(\omega)|^2 e^{-i\omega h} d\omega \\
&= \sum_{j=h}^{\infty} \psi_j \bar{\psi}_{j-h}.
\end{aligned}$$

Il en résulte que les p.a. $(z_t; t \in \mathbb{Z})$ et $(z'_t; t \in \mathbb{Z})$ ont la même fonction d'autocovariance. Les coefficients de leur décomposition de Wold sont donc égaux et leurs innovations respectives e_t et e'_t ont même variance σ^2 . Considérons la variance de l'erreur de prévision de z'_t faite à partir d'une combinaison linéaire d'un nombre arbitraire m de valeurs passées $z'_{t-1}, \dots, z'_{t-m}$:

$$\text{var}\left(z'_t - \sum_{h=1}^n \pi_h z'_{t-h}\right). \quad (4.6)$$

Par application de la Proposition 2.38, l'infimum de (4.6) est atteint et vaut

$$\text{var}\left[z'_t - P(z'_t | V_{t-1})\right] = \text{var}(e'_t) = \sigma^2. \quad (4.7)$$

Or (4.6) s'écrit

$$\text{var}\left(\sum_{j=0}^{\infty} \psi'_j e'_{t-j} - \sum_{h=1}^n \pi_h \sum_{j=0}^{\infty} \psi'_j e'_{t-h-j}\right) \geq |\psi'_0|^2 \text{var}(e'_t)$$

ce qui vaut $|\psi'_0|^2 \sigma^2$, en vertu du théorème de Pythagore (Proposition 2.33). Or $|\psi'_0|^2 = 1/|b|^2 > 1$ donc (4.6) est supérieur ou égal à un nombre strictement supérieur à σ^2 ce qui contredit (4.7). \square

Remarque. On constate que la démonstration ci-dessus, inspirée par Doob (1953), utilise quelques éléments de l'approche spectrale. Notons que la démonstration purement géométrique de Grenander et Rosenblatt (1957, p. 71) est fautive.

Proposition 4.9. 4.8'. *Une condition nécessaire et suffisante pour qu'un processus ARMA stationnaire $(z_t; t \in \mathbb{Z})$, tel que les polynômes $\phi_p(B)$ et $\theta_q(B)$ n'ont pas de zéro en commun, soit inversible est que les zéros de $\theta_q(B)$ soient supérieurs à 1 en module.*

Démonstration. Elle est divisée en quatre parties.

(a) Considérons d'abord un p.s.s. à coefficients complexes satisfaisant

$$\sum_{j=0}^{\infty} \varphi_j z_{t-j} = e_t - \mu e_{t-1},$$

où $\varphi_0 = 1$ et la série du membre de gauche, qu'on note u_t , converge en moyenne. Montrons que le processus est inversible si $|\mu| < 1$. Remarquons d'abord que $V_2(u;t) = V_2(z;t)$ pour tout t et que les e_t , qui sont les innovations de $(z_t; t \in \mathbb{Z})$ par hypothèse sont également les innovations de $(u_t; t \in \mathbb{Z})$.

Appliquons la Proposition 3.14 au cas scalaire. Si $|\mu| < 1$, alors $\lim \mu^m = 0$ pour $m \rightarrow \infty$ et donc

$$e_t = \sum_{m=0}^{\infty} \mu^m u_{t-m}.$$

(b) Considérons maintenant un p.s.s. à coefficients complexes $(z_t; t \in \mathbb{Z})$ satisfaisant

$$\sum_{j=0}^p \varphi_j z_{t-j} = b_t - \mu b_{t-1}, \quad (4.8)$$

où p est fini, et tel que μ^{-1} n'est pas un zéro de $\sum_{j=0}^p \varphi_j B^j$. Notons que $(b_t; t \in \mathbb{Z})$ n'est pas nécessairement un bruit blanc. On suppose qu'il existe des coefficients π_j tels que

$$b_t = \sum_{j=0}^{\infty} \pi_j z_{t-j} \quad (4.9)$$

et on veut montrer que cela implique que $|\mu| < 1$. Remplaçons b_t dans (4.8). En égalant les coefficients de z_{t-j} , il vient

$$\pi_0 = 1, \quad \pi_j - \mu \pi_{j-1} = \varphi_j \quad (j \geq 1)$$

d'où

$$\pi_j = \sum_{h=0}^j \mu^{j-h} \varphi_h = \mu^j \sum_{h=0}^j \mu^{-h} \varphi_h.$$

Par conséquent, la série

$$\sum_{j=0}^{\infty} \mu^j \left(\sum_{h=0}^j \mu^{-h} \varphi_h \right) z_{t-j}$$

converge en moyenne, donc également

$$\sum_{j=p}^{\infty} \mu^j \left(\sum_{h=0}^j \mu^{-h} \varphi_h \right) z_{t-j}$$

qui en diffère par un nombre fini de termes. En mettant en évidence le facteur commun, non nul par hypothèse, il vient que la série $\sum_j \mu^j z_{t-j}$ converge en moyenne, d'où $|\mu|^{2j} \text{var}(z_{t-j}) = |\mu|^{2j} \gamma_0 \rightarrow 0$ pour $j \rightarrow \infty$ et donc $|\mu| < 1$.

(c) La condition suffisante s'établit de la façon suivante. Désignons par μ_j^{-1} les zéros de $\theta_q(B)$ de sorte que

$$\theta_q(B) = \prod_{j=1}^q (1 - \mu_j B). \quad (4.10)$$

On introduit les p.a. à coefficients complexes $(b_t^{(\ell)}; t \in \mathbb{Z})$ définis par

$$b_t^{(\ell)} = e_t - \theta_1^{(\ell)} e_{t-1} - \dots - \theta_\ell^{(\ell)} e_{t-\ell},$$

où les coefficients $\theta_i^{(\ell)}$ sont tels que

$$1 - \theta_1^{(\ell)} B - \dots - \theta_\ell^{(\ell)} B^\ell = \prod_{j=1}^{\ell} (1 - \mu_j B),$$

pour $\ell = 1, 2, \dots, q$. On a

$$\begin{aligned} b_t^{(q)} &= b_t^{(q-1)} - \mu_q b_{t-1}^{(q-1)} \\ b_t^{(q-1)} &= b_t^{(q-2)} - \mu_{q-1} b_{t-1}^{(q-2)} \\ &\dots \\ b_t^{(1)} &= e_t - \mu_1 e_{t-1} \end{aligned}$$

et $b_t^{(q)} = z_t - \phi_1 z_{t-1} - \dots - \phi_p z_{t-p}$. Puisque $|\mu_i| < 1$ ($i = 1, \dots, q$), on peut appliquer (a) pour exprimer successivement $b_t^{(q-1)}$, $b_t^{(q-2)}$, ..., $b_t^{(1)}$, e_t comme séries en les z_{t-i} , convergeant en moyenne.

(d) La condition nécessaire peut se démontrer comme suit. Considérons un facteur quelconque de (4.10) et définissons

$$b_t^{(\ell)} = \prod_{\substack{j=1 \\ j \neq \ell}}^q (1 - \mu_j B) e_t.$$

Il s'ensuit que

$$b_t^{(\ell)} - \mu_\ell b_{t-1}^{(\ell)} = \sum_{j=0}^p \varphi_j z_{t-j}$$

avec $\varphi_0 = 1$, $\varphi_j = -\phi_j$ ($j = 1, \dots, p$). Par hypothèse, e_t s'exprime comme série convergente en moyenne en les z_{t-i} , donc également $b_t^{(\ell)}$. Par application de (b), cela entraîne donc, pourvu que μ_ℓ^{-1} ne soit pas un zéro de $\phi_p(B)$, que $|\mu_\ell| < 1$. \square

4.3. Caractérisation des processus ARMA

Les caractérisations des processus purs AR et MA résultent directement de la particularisation des Propositions 3.8 et 3.11. Il est également possible de caractériser les processus ARMA multivariés.

Notons d'abord que pour un processus ARMA(p, q) stationnaire, (3.20) devient, en posant $\gamma_h'' = \mathbf{\Gamma}_{t, t-h}''$:

$$\boldsymbol{\gamma}_h = \boldsymbol{\gamma}'_h + \sum_{i=1}^p \boldsymbol{\phi}_i \boldsymbol{\gamma}_{h-i} \quad (4.11)$$

où $\boldsymbol{\gamma}'_h = \mathbf{0}$ pour $|h| > q$. Soient $P > p$ et $Q > q$. On peut écrire un système de $P+1$ équations obtenu en prenant (4.11) pour $h = Q, \dots, Q+P$:

$$\begin{bmatrix} \boldsymbol{\gamma}_Q \\ \dots \\ \boldsymbol{\gamma}_{Q+P-1} \\ \boldsymbol{\gamma}_{Q+P} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\gamma}_{Q-1} & \dots & \boldsymbol{\gamma}_{Q-P+1} & \boldsymbol{\gamma}_{Q-P} \\ \dots & \dots & \dots & \dots \\ \boldsymbol{\gamma}_{Q+P-2} & \dots & \boldsymbol{\gamma}_Q & \boldsymbol{\gamma}_{Q-1} \\ \boldsymbol{\gamma}_{Q+P-1} & \dots & \boldsymbol{\gamma}_{Q+1} & \boldsymbol{\gamma}_Q \end{bmatrix} \begin{bmatrix} \boldsymbol{\phi}_1 \\ \boldsymbol{\phi}_2 \\ \dots \\ \boldsymbol{\phi}_P \end{bmatrix} \quad (4.12)$$

où $\boldsymbol{\phi}_{p+1} = \dots = \boldsymbol{\phi}_P = \mathbf{0}$. Considérons le système de $Pk+1$ équations linéaires obtenu en prenant les P premiers blocs de lignes de (4.12) ainsi que la i ème ligne du $(P+1)$ ème bloc, et en se limitant à la j ème colonne du premier membre et donc à la j ème colonne $\boldsymbol{\phi}_j(P)$ de $\boldsymbol{\phi}(P) = (\boldsymbol{\phi}_1^\top, \boldsymbol{\phi}_2^\top, \dots, \boldsymbol{\phi}_p^\top)^\top$. Notons

$$\mathbf{A}(P, Q) = \begin{bmatrix} \boldsymbol{\gamma}_{Q-1} & \dots & \boldsymbol{\gamma}_{Q-P+1} & \boldsymbol{\gamma}_{Q-P} \\ \dots & \dots & \dots & \dots \\ \boldsymbol{\gamma}_{Q+P-2} & \dots & \boldsymbol{\gamma}_Q & \boldsymbol{\gamma}_{Q-1} \end{bmatrix}, \quad \mathbf{C}(P, Q) = \begin{bmatrix} \boldsymbol{\gamma}_Q \\ \dots \\ \boldsymbol{\gamma}_{Q+P-1} \end{bmatrix},$$

$\mathbf{R}(P, Q) = [\boldsymbol{\gamma}_{Q+P-1}, \dots, \boldsymbol{\gamma}_Q]$. Notons $\mathbf{R}_i(P, Q)$ la i ème ligne de $\mathbf{R}(P, Q)$, $\mathbf{C}_j(P, Q)$ la j ème colonne de $\mathbf{C}(P, Q)$, et $e_{ij}(P, Q)$ l'élément (i, j) de $\boldsymbol{\gamma}_{Q+P}$. Puisque le système

$$\begin{bmatrix} \mathbf{C}_j(P, Q) \\ e_{ij}(P, Q) \end{bmatrix} = \begin{bmatrix} \mathbf{A}(P, Q) \\ \mathbf{R}_i(P, Q) \end{bmatrix} \boldsymbol{\phi}_j(P)$$

est compatible, on a

$$D_{ij}(P, Q) = \det \begin{bmatrix} \mathbf{A}(P, Q) & \mathbf{C}_j(P, Q) \\ \mathbf{R}_i(P, Q) & e_{ij}(P, Q) \end{bmatrix} = 0 \quad (4.13)$$

pour $i, j = 1, \dots, k$, et donc la matrice $\mathbf{D}(P, Q)$ constituée des $D_{ij}(P, Q)$ est nulle, pour $P > p$ et $Q > q$ (Tiao et Box, 1981).

Les $D_{ij}(P, Q)$ peuvent être calculés successivement en appliquant les règles relatives aux matrices partitionnées de la forme

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}.$$

Ces règles sont

$$\left. \begin{aligned} \det(\mathbf{A}) &= \det(\mathbf{A}_{11}) \det(\mathbf{A}_{22} - \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{A}_{12}), \\ \mathbf{A}^{-1} &= \begin{bmatrix} \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1} \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{A}_{21} \mathbf{A}_{11}^{-1} & -\mathbf{A}_{11}^{-1} \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \\ -\mathbf{A}_{22}^{-1} \mathbf{A}_{21} \mathbf{A}_{11}^{-1} & \mathbf{A}_{22}^{-1} \end{bmatrix}, \end{aligned} \right\} \quad (4.14)$$

avec $\mathbf{A}^{22} = (\mathbf{A}_{22} - \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{A}_{12})^{-1}$. Si l'on considère la matrice partitionnée

$$\mathbf{A}(P, Q) = \begin{bmatrix} \mathbf{A}(P-1, Q) & \mathbf{C}(P-1, Q-P) \\ \mathbf{R}(P-1, Q) & \gamma_{Q-1} \end{bmatrix}$$

on peut déterminer $\det[\mathbf{A}(P, Q)]$ et $(\mathbf{A}(P, Q))^{-1}$ dès qu'on connaît $\det[\mathbf{A}(P-1, Q)]$ et $(\mathbf{A}(P-1, Q))^{-1}$. En appliquant à nouveau (4.13) à la partition implicite dans (4.13), on est en mesure d'évaluer le déterminant. Cet algorithme a été suggéré par de Falguerolles (1980) dans le cas scalaire.

La caractérisation (4.13) est à la base de la méthode du coin (Beguin, Gouriéroux et Monfort, 1980, et Lindberger, 1973).

Considérons les P premières lignes de (4.12) pour $Q = 0$ et résolvons le système par rapport à ϕ_p . En envisageant cette solution comme fonction de P on obtient la fonction matricielle des coefficients de régression partielle (Huyberegts, 1983, où cette appellation est justifiée) aussi connue sous le nom de fonction matricielle d'autorégression partielle (Tiao et Box, 1981) qui généralise la fonction d'autocorrélation partielle du cas univarié. L'algorithme est alors la généralisation due à Robinson (1967) de l'algorithme de Levinson (1949) et de Durbin (1960).

4.4. Exemples de processus ARMA

Proposition 4.10. Soit $(z_t; t \in \mathbb{Z})$ un p.a.s. qui vérifie

$$z_t = \sum_{j=0}^q y_{t-j,j} b_{t-j}$$

où $(b_t; t \in \mathbb{Z})$ est un b.b.c. de variance σ^2 et $(\mathbf{y}_t; t \in \mathbb{Z})$, avec $\mathbf{y}_t = (y_{t,0}, \dots, y_{t-q,q})^\top$, est un p.s.s.p.i. vectoriel de dimension $q+1$, tels que ces deux p.a. sont indépendants. Le p.a.s. $(z_t; t \in \mathbb{Z})$ est alors un p.s.s. MA(q).

Démonstration. Notons $\gamma_h^* = \langle \mathbf{y}_t, \mathbf{y}_{t-h} \rangle$. Il vient directement, pour $h > 0$:

$$\begin{aligned} \langle z_t, z_{t-h} \rangle &= \left\langle \sum_{j=0}^q y_{t-j,j} b_{t-j}, \sum_{i=0}^q y_{t-h-i,i} b_{t-h-i} \right\rangle \\ &= \sum_{j=0}^q \sum_{i=0}^q \langle y_{t-j,j} b_{t-j}, y_{t-h-i,i} b_{t-h-i} \rangle \\ &= \sum_{j=0}^q \sum_{i=0}^q E(y_{t-j,j} b_{t-j} y_{t-h-i,i} b_{t-h-i}). \end{aligned}$$

Compte tenu de l'hypothèse d'indépendance (notée par le symbole $\perp\!\!\!\perp$) et donc de ce que

$$b_{t-j} \perp\!\!\!\perp y_{t-j,j} y_{t-h-i,i} b_{t-h-i} \text{ si } h+i \neq j,$$

il demeure

$$\begin{aligned} \langle z_t, z_{t-h} \rangle &= \sum_{j=h}^q E[y_{t-j,j} b_{t-j}^2 y_{t-j,j-h}] \\ &= \begin{cases} \sigma^2 \sum_{j=h}^q (\gamma_0^*)_{j,j-h} & \text{si } 0 \leq h \leq q \\ 0 & \text{si } h > q. \end{cases} \end{aligned}$$

□

Proposition 4.11. Soit $(z_t; t \in \mathbb{Z})$ un p.a.s. qui vérifie

$$z_t = \sum_{j=0}^q h_j(\mathbf{b}_{t-j-1}) b_{t-j}$$

où $\mathbf{b}_s = (b_s, \dots, b_{s-r+1})^\top$, $(b_t; t \in \mathbb{Z})$ est un b.b.c. de variance σ^2 et les $h_j(\cdot)$, des fonctions telles que $(\mathbf{h}_t; t \in \mathbb{Z})$ est un p.s.s. avec $\mathbf{h}_t = (h_1(\mathbf{b}_{t-2}), \dots, h_q(\mathbf{b}_{t-q-1}))$. Alors le p. a. $(z_t; t \in \mathbb{Z})$ est un p.a.s. MA(q).

Démonstration. De façon similaire à la démonstration de la Proposition 4.10, il vient

$$\langle z_t, z_{t-h} \rangle = \sum_{j=0}^q \sum_{i=0}^q E[h_j(\mathbf{b}_{t-j-1}) b_{t-j} h_i(\mathbf{b}_{t-h-i-1}) b_{t-h-i}].$$

Puisque $(b_t; t \in \mathbb{Z})$ est un b.b., on a

$$\begin{aligned} b_{t-j} &\perp\!\!\!\perp h_j(\mathbf{b}_{t-j-1}) h_i(\mathbf{b}_{t-h-i-1}) b_{t-h-i} \quad \text{si } h+i > j, \\ b_{t-h-i} &\perp\!\!\!\perp h_j(\mathbf{b}_{t-j-1}) h_i(\mathbf{b}_{t-h-i-1}) b_{t-j} \quad \text{si } h+i < j. \end{aligned}$$

Par conséquent, pour $h > 0$,

$$\begin{aligned} \langle z_t, z_{t-h} \rangle &= \sum_{j=h}^q E[h_j(\mathbf{b}_{t-j-1}) h_{j-h}(\mathbf{b}_{t-j-1}) b_{t-j}^2] \\ &= \begin{cases} \sigma^2 \sum_{j=h}^q E[h_j(\mathbf{b}_{t-j-1}) h_{j-h}(\mathbf{b}_{t-j-1})] & \text{si } 0 \leq h \leq q \\ 0 & \text{si } h > q. \end{cases} \end{aligned}$$

ce qui établit la proposition. □

Remarque. On reconnaît les processus dépendant de l'état ("state dependent processes") introduits par Priestley (1980).

4.5. La fonction d'autocovariance

La relation (4.11) peut se réécrire

$$\gamma_h - \sum_{i=1}^p \phi_i \gamma_{h-i} = \gamma'_h$$

où γ'_h est l'autocovariance d'ordre h du processus constitué à partir de l'opérateur MA. Puisque $\gamma_{-h} = \gamma'_h$, on peut écrire un système d'équations en les éléments de $\gamma_0, \gamma_1, \dots, \gamma_p$, au moyen de (4.11) pour $h = 0, \dots, p$. Ce système permet de déterminer les $k(k+1)/2 + pk^2$ inconnues contenues dans ces $p+1$ matrices. On peut vérifier que ce système comporte autant d'équations que d'inconnues. Un algorithme pour constituer ce système a été présenté par Ansley (1980) et par Nicholls et Hall (1979) généralisant McLeod (1975). Il reste alors à résoudre ce système. Kohn et Ansley (1982) présentent un algorithme amélioré.

Dans le cas univarié, un algorithme plus efficace a été proposé par Tunncliffe Wilson (1979). Il fait l'objet de cette section. Considérons le processus ARMA(p, q) univarié défini par l'équation $\phi(B)z_t = \theta(B)e_t$. On veut déterminer γ_h pour $h = 0, \dots, H$. Commençons par introduire une notation qui sera fréquemment utilisée. Si $\mu_r(B)$ est un polynôme de degré inférieur ou égal à r , on note $\mu_r^*(B) = B^r \mu_r(B^{-1})$, ce qui représente un polynôme de degré inférieur ou égal à r . Remarquons que le coefficient de B^r dans $\mu_r(B)$ vaut $\mu_r^*(0)$ tandis que le terme indépendant de $\mu_r(B)$ vaut le coefficient de B^r de $\mu_r^*(B)$. De plus, si $b \neq 0$ est un zéro de $\mu_r(B)$, alors b^{-1} est un zéro de $\mu_r^*(B)$. Notons D le disque unité : $D = \{B \in \mathbb{C}; |B| < 1\}$.

L'algorithme de Tunncliffe Wilson est constitué des six étapes suivantes, où l'on note $m = \max(p, q)$.

Étape 1: On pose

$$\begin{aligned}\beta_m(B) &= \theta(B)\theta(B^{-1}) = \sum_{h=-q}^q \beta_h B^h, \\ b_m(B) &= \frac{1}{2}\beta_0 + \sum_{h=1}^q \beta_h B^h, \\ \phi_m(B) &= \phi(B), \\ v_0(B) &= \frac{1}{2}\beta_0.\end{aligned}$$

Étape 2: Pour $h = m-1, \dots, 0$, on calcule

$$\begin{aligned}\alpha_{h+1} &= \phi_{h+1}^*(0), \\ \phi_h(B) &= [\phi_{h+1}(B) - \alpha_{h+1}\phi_{h+1}^*(B)](1 - \alpha_{h+1}^2)^{-1},\end{aligned}$$

pourvu que $|\alpha_{h+1}| \neq 1$.

Étape 3: Pour $h = m-1, \dots, 0$, on calcule

$$\begin{aligned}\delta_{h+1} &= b_{h+1}^*(0), \\ b_h(B) &= b_{h+1}(B) - \delta_{h+1}B\phi_h^*(B), \\ \beta_h &= b_h(B) + b_h^*(B).\end{aligned}$$

Étape 4: Pour $h = 1, \dots, m-1$, on calcule

$$v_{h+1}(B) = \frac{v_h(B) - \alpha_{h+1} B v_h^*(B) + \delta_{h+1} B^{h+1} - \delta_{h+1} \alpha_{h+1}}{1 - \alpha_{h+1}^2}.$$

Etape 5: On effectue $g(B) = v_m(B)/\phi_m(B)$.

Etape 6: On obtient finalement

$$\gamma_0 = 2g_0\sigma^2,$$

$$\gamma_{-h} = \gamma_h = g_h\sigma^2 \text{ pour } h = 1, \dots, H.$$

Afin de justifier cet algorithme, on établit d'abord une suite de lemmes.

Lemme 4.12.1 *Pour tout h , $0 \leq h \leq m$, pour autant que $|\alpha_{h+1}| \neq 1$, $\phi_h(B)$ est un polynôme de degré inférieur ou égal à h et son terme indépendant vaut 1.*

Démonstration. La proposition est vraie lorsque $h = m$. Supposons qu'elle soit vraie pour $\phi_{h+1}(B)$ et montrons qu'elle est vraie pour $\phi_h(B)$. En effet, $\phi_{h+1}^*(B)$ est un polynôme de degré au plus égal à $h+1$ et le coefficient de B^{h+1} vaut α_{h+1} . Dès lors, $\phi_h(B)$ est de degré au plus égal à $h+1$. Mais on vérifie aisément que le terme en B^{h+1} vaut zéro. De plus,

$$\phi_h(0) = \frac{\phi_{h+1}(0) - \alpha_{h+1}\phi_{h+1}^*(0)}{1 - \alpha_{h+1}^2} = 1.$$

Lemme 4.12.2 *Si $p < q = m$, $\phi_p(B) = \phi(B)$ et $\alpha_{h+1} = 0$ pour $h = p, \dots, m-1$.*

Démonstration. En effet $\phi_m(B) = \phi(B)$ est de degré p . Les polynômes $\phi_{m-1}(B), \dots, \phi_p(B)$ sont donc de degré inférieur ou égal à p et donc $\alpha_m = \alpha_{m-1} = \dots = \alpha_{p+1} = 0$. Le lemme résulte alors de la définition de $\phi_h(B)$. \square

Lemme 4.12.3 *Pour tout h , $0 \leq h < m$, tel que $|\alpha_{h+1}| \neq 1$, on a $\phi_{h+1}(B) = \phi_h(B) + \alpha_{h+1} B \phi_h^*(B)$.*

Démonstration. Par définition

$$\phi_{h+1}(B) - \alpha_{h+1} \phi_{h+1}^*(B) = (1 - \alpha_{h+1}^2) \phi_h(B).$$

En remplaçant B par B^{-1} et multipliant les deux membres par $\alpha_{h+1} B^{h+1}$, il vient

$$\alpha_{h+1} \phi_{h+1}^*(B) - \alpha_{h+1}^2 \phi_{h+1}(B) = (1 - \alpha_{h+1}^2) \alpha_{h+1} B \phi_h^*(B).$$

Par addition membre à membre des deux équations et simplification par $(1 - \alpha_{h+1}^2)$ on trouve la relation annoncée. \square

Lemme 4.12.4 *Si $(z_i; t \in \mathbb{Z})$ est un p.s.s., alors pour tout $h = 1, \dots, p$, $\phi_h(B)$ est un polynôme dont les zéros sont à l'extérieur du disque unité D et $|\alpha_h| < 1$.*

Démonstration. Par l'hypothèse de stationnarité et compte tenu du Lemme 4.12.2, tous les zéros de $\phi_p(B) = \phi(B)$ sont à l'extérieur de D . Puisque $\phi_p(0) = 1$, le coefficient de B^p est le produit des inverses des zéros du polynôme, d'où $|\alpha_h| < 1$. Montrons que si la proposition est vraie pour $\phi_{h+1}(B)$, elle l'est aussi pour $\phi_h(B)$.

Supposons d'abord que $|\alpha_h| = 1$. On peut alors écrire, en vertu des Lemmes 4.12.1 et 4.12.2 :

$$\phi_h(B) = \pm \prod_{j=1}^s (B + \mu_j), \quad \phi_h^*(B) = \pm B^{h-s} \prod_{j=1}^s (1 + \mu_j B) \text{ avec } \prod_{j=1}^s \mu_j = 1, \quad (4.-6)$$

où $s \leq h$. Par application du Lemme 4.12.3, on a

$$\phi_{h+1}(B) = f_1(B) + g_1(B) \quad \text{où} \quad f_1(B) = \phi_h(B) \quad \text{et} \quad g_1(B) = \alpha_{h+1} B \phi_h^*(B).$$

Deux cas sont possibles, compte tenu de (4.13) :

- (i) tous les zéros de $\phi_h(B)$ sont situés sur la frontière de D , ce sont aussi des zéros de $\phi_h^*(B)$ donc de $\phi_{h+1}(B)$ ce qui contredit l'hypothèse de récurrence;
- (ii) $\phi_h(B)$ a au moins un zéro à l'intérieur de D . Sachant que $|\alpha_{h+1}| < 1$, il vient pour $|B| = 1$:

$$|g_1(B)| = |\alpha_{h+1}| \left| \prod_{j=1}^s (1 + \mu_j B) \right| < \left| \prod_{j=1}^s (B + \mu_j) \right| = |f_1(B)|.$$

Le théorème de Rouché permet alors d'affirmer que $f_1(B)$ et $f_1(B) + g_1(B)$ ont le même nombre de zéros à l'intérieur de D . Or $\phi_h(B)$ en a au moins un, donc aussi $\phi_{h+1}(B)$, ce qui contredit l'hypothèse de récurrence.

On a donc établi que $|\alpha_h| \neq 1$. Montrons maintenant que $\phi_h(B)$ a tous ses zéros à l'extérieur de D . Considérons les polynômes

$$f_2(B) = (1 - \alpha_{h+1}^2)^{-1} \phi_{h+1}(B), \quad g_2(B) = -(1 - \alpha_{h+1}^2)^{-1} \alpha_{h+1} \phi_{h+1}^*(B).$$

Pour tout complexe B de module 1 :

$$\begin{aligned} |g_2(B)| &= |1 - \alpha_{h+1}^2|^{-1} |\alpha_{h+1}| |\phi_{h+1}^*(B)| \\ &= |1 - \alpha_{h+1}^2|^{-1} |\alpha_{h+1}| |\phi_{h+1}(B)| < |f_2(B)|, \end{aligned}$$

car $|\alpha_{h+1}| < 1$ par l'hypothèse de récurrence. Le théorème de Rouché permet donc d'affirmer que $f_2(B)$ et $f_2(B) + g_2(B) = \phi_h(B)$ ont le même nombre de zéros à l'intérieur de D . Or, par l'hypothèse de récurrence, $\phi_{h+1}(B)$ n'en a pas. Il en est donc de même de $\phi_h(B)$. Soit b un zéro de $\phi_h(B)$ de module 1. Grâce au Lemme 4.12.3 on vérifie que $\phi_{h+1}(b) = 0$, d'où une contradiction.

Il reste à montrer que $|\alpha_h| < 1$. Considérons les polynômes

$$f_3(B) = (1 - \alpha_h^2)^{-1} \alpha_h \phi_h^*(B), \quad g_3(B) = (1 - \alpha_h^2)^{-1} \phi_h(B).$$

Supposons $|\alpha_h| > 1$. On aurait donc pour tout complexe B de module 1

$$\begin{aligned} |g_3(B)| &= |1 - \alpha_h^2|^{-1} |\phi_h(B)| = |1 - \alpha_h^2|^{-1} |\phi_h^*(B)| \\ &< |1 - \alpha_h^2|^{-1} |\alpha_h| |\phi_h^*(B)| = |f_3(B)|. \end{aligned}$$

Le théorème de Rouché permet alors d'affirmer que $f_3(B)$ et $f_3(B) + g_3(B) = \phi_{h-1}(B)$ ont le même nombre de zéros à l'intérieur de D . Or $|\alpha_h| \neq 0$ par hypothèse, donc les zéros de $\phi_h^*(B)$ sont non nuls, $\phi_h(B)$ est de degré h et les h zéros de $\phi_h(B)$ sont les inverses de ceux de $\phi_h^*(B)$. Le nombre de zéros de $\phi_h^*(B)$ à l'intérieur de D , qui vaut h en vertu de ce qui précède, est donc égal au nombre de zéros de $\phi_{h-1}(B)$ à l'intérieur de D . Puisque $\phi_{h-1}(B)$ est au plus de degré $h-1$, il y a contradiction. \square

Remarque. Ce lemme justifie la division par $1 - \alpha_{h+1}^2$ dans les étapes 2 et 4 de l'algorithme, sous l'hypothèse de stationnarité.

Lemme 4.12.5 *Si $|\alpha_h| < 1$ pour $h = 1, \dots, p$, alors les zéros de $\phi(B)$ sont à l'extérieur du disque unité D .*

Démonstration. Considérons à nouveau les polynômes $f_1(B)$ et $g_1(B)$ introduits dans la démonstration du Lemme 4.12.4. Il vient pour $|B| = 1$:

$$|g_1(B)| = |\alpha_{h+1}| |B| |\phi_h^*(B)| = |\alpha_{h+1}| |\phi_h(B)| < |\phi_h(B)| = |f_1(B)|.$$

Le théorème de Rouché permet donc d'affirmer que $f_1(B) = \phi_h(B)$ et $f_1(B) + g_1(B) = \phi_{h+1}(B)$ ont le même nombre de zéros à l'intérieur de D . Donc $\phi_p(B)$, $\phi_{p-1}(B)$, ..., $\phi_1(B)$ et $\phi_0(B)$ ont le même nombre de zéros à l'intérieur de D . Puisque $\phi_0(B)$ est une constante, $\phi_p(B)$ n'a aucun zéro à l'intérieur de D . On peut vérifier que si b est un zéro de $\phi_p(B)$ de module 1, alors $\phi_{p-1}(b) = \dots = \phi_0(b) = 0$, d'où une contradiction.

Remarque. Ce lemme permet de vérifier l'hypothèse de stationnarité sans la détermination des zéros de $\phi(B)$. Voir Åström (1970, p. 119) et Anderson (1975).

Lemme 4.12.6 *Pour tout h , $0 \leq h \leq m$, $b_h(B)$ est un polynôme de degré inférieur ou égal à h et son terme indépendant vaut $\beta_0/2$.*

Démonstration. La proposition est vraie lorsque $h = m$. Supposons qu'elle soit vraie pour $b_{h+1}(B)$ et montrons qu'elle est vraie pour $b_h(B)$. Par le Lemme 4.11.1, $B\phi_h^*(B)$ est un polynôme de degré $h+1$, dont le terme en B^{h+1} a 1 pour coefficient. Il s'ensuit que $b_h(B)$ est un polynôme de degré inférieur ou égal à $h+1$. De plus, le coefficient du terme en B^{h+1} de $b_{h+1}(B)$ valant δ_{h+1} par définition, le terme en B^{h+1} de $b_h(B)$ vaut 0. Finalement, $b_h(0) = b_{h+1}(0) = \beta_0/2$. \square

Lemme 4.12.7 *Si $q < p = m$, $b_h(B) = b_q(B)$ et $\delta_{h+1} = 0$ pour $h = q, q+1, \dots, m-1$.*

Démonstration. En effet, $b_m(B) = b_q(B)$ est de degré q . Les polynômes $b_{m-1}(B)$, ..., $b_q(B)$ sont donc de degré inférieur ou égal à q et donc $\delta_m = \delta_{m-1} = \dots = \delta_{q+1} = 0$. Le lemme résulte alors de la définition de $b_h(B)$. \square

Lemme 4.12.8 *Pour tout h , $0 \leq h < m$:*

$$v_{h+1}(B) + \alpha_{h+1} v_{h+1}^*(B) = v_h(B) + \delta_{h+1} B^{h+1}.$$

Démonstration. Par définition

$$v_{h+1}(B) = \frac{v_h(B) - \alpha_{h+1}Bv_h^*(B) + \delta_{h+1}B^{h+1} - \delta_{h+1}\alpha_{h+1}}{1 - \alpha_{h+1}^2}.$$

En remplaçant B par B^{-1} et en multipliant les deux membres par $\alpha_{h+1}B^{h+1}$, il vient

$$\alpha_{h+1}v_{h+1}^*(B) = \frac{\alpha_{h+1}Bv_h^*(B) - \alpha_{h+1}^2v_h(B) + \delta_{h+1}\alpha_{h+1} - \alpha_{h+1}^2\delta_{h+1}B^{h+1}}{1 - \alpha_{h+1}^2}.$$

Par addition membre à membre des deux équations et simplification par $1 - \alpha_{h+1}^2$, on trouve la relation annoncée. \square

Lemme 4.12.9 Pour tout h , $0 \leq h \leq m$:

$$\beta_h(B) = \phi_h(B^{-1})v_h(B) + \phi_h(B)v_h(B^{-1}).$$

Démonstration. La proposition est vraie pour $h = 0$ car

$$\beta_0(B) = b_0(B) + b_0(B^{-1}) = \beta_0$$

par le Lemme 4.12.6, tandis que

$$\phi_0(B^{-1})v_0(B) + \phi_0(B)v_0(B^{-1}) = \beta_0$$

par le Lemme 4.12.1 et la définition de $v_0(B)$. Montrons que si la proposition est vraie pour $\beta_h(B)$, elle l'est aussi pour $\beta_{h+1}(B)$. Or

$$\begin{aligned} \beta_{h+1}(B) &= b_{h+1}(B) + b_{h+1}(B^{-1}) \\ &= b_h(B) + \delta_{h+1}B\phi_h^*(B) + b_h(B^{-1}) + \delta_{h+1}B^{-1}\phi_h^*(B^{-1}) \\ &= \beta_h(B) + \delta_{h+1}B^{h+1}\phi_h(B^{-1}) + \delta_{h+1}B^{-h-1}\phi_h(B) \\ &= \phi_h(B^{-1})[v_h(B) + \delta_{h+1}B^{h+1}] + \phi_h(B)[v_h(B^{-1}) + \delta_{h+1}B^{-h-1}] \end{aligned}$$

par l'hypothèse de récurrence. Grâce au Lemme 4.12.8, il vient

$$\begin{aligned} \beta_{h+1}(B) &= \phi_h(B^{-1})[v_{h+1}(B) + \alpha_{h+1}v_{h+1}^*(B)] + \phi_h(B)[v_{h+1}(B^{-1}) + \alpha_{h+1}v_{h+1}^*(B^{-1})] \\ &= \phi_h(B^{-1})v_{h+1}(B) + \phi_h(B)v_{h+1}(B^{-1}) \\ &\quad + \alpha_{h+1}[B\phi_{h+1}^*(B)v_{h+1}(B^{-1}) + B^{-1}\phi_h^*(B^{-1})v_{h+1}(B)] \\ &= \phi_{h+1}(B^{-1})v_{h+1}(B) + \phi_{h+1}(B)v_{h+1}(B^{-1}) \end{aligned}$$

par application du Lemme 4.12.3, ce qui termine la démonstration. \square

Proposition 4.12. Les γ_h , $h = 0, \dots, H$, déterminés par l'algorithme de Tunnicliffe Wilson sont les autocovariances du processus ARMA(p, q) défini par (4.1).

Démonstration. Considérons $\gamma(B)$, la fonction génératrice associée à $(\gamma_h; h \in \mathbb{Z})$,

$$\gamma(B) = \sum_{h=-\infty}^{\infty} \gamma_h B^h$$

considérée comme série formelle en B . Compte tenu des étapes 5 et 6 de l'algorithme, il vient

$$\begin{aligned} \gamma(B) &= \sigma^2 \left[2\gamma_0 + \sum_{h=1}^{\infty} g_h B^h + \sum_{h=-1}^{-\infty} g_h B^h \right] \\ &= \sigma^2 [g(B) + g(B^{-1})] \\ &= \sigma^2 \left[\frac{v_m(B)}{\phi_m(B)} + \frac{v_m(B^{-1})}{\phi_m(B^{-1})} \right] \\ &= \sigma^2 \frac{\phi_m(B^{-1})v_m(B) + \phi_m(B)v_m(B^{-1})}{\phi_m(B)\phi_m(B^{-1})} \\ &= \sigma^2 \frac{\beta_m(B)}{\phi_m(B)\phi_m(B^{-1})} \\ &= \sigma^2 \frac{\theta_q(B)\theta_q(B^{-1})}{\phi_m(B)\phi_m(B^{-1})}, \end{aligned}$$

grâce aux Lemmes 4.12.9, 4.12.2 et 4.12.7. On peut donc écrire

$$\gamma(B) = \sigma^2 \psi(B)\psi(B^{-1}) \quad \text{où} \quad \psi(B) = \phi_p^{-1}(B)\theta_q(B).$$

Or

$$\begin{aligned} \text{cov}(z_t, z_{t-h}) &= \text{cov}(\psi(B)e_t, \psi(B)e_{t-h}) \\ &= \left(\sum_{j=0}^{\infty} \psi_j \psi_{j+h} \right) \sigma^2 \end{aligned}$$

pour $h \geq 0$. Si l'on note $\psi_j = 0$ pour $j < 0$, ceci est également valable lorsque $h < 0$. Considérons

$$\begin{aligned} \sum_{h=-\infty}^{\infty} \text{cov}(z_t, z_{t-h}) B^h &= \sigma^2 \sum_{h=-\infty}^{\infty} \left(\sum_{j=0}^{\infty} \psi_j \psi_{j+h} \right) B^h \\ &= \sigma^2 \sum_{h=-\infty}^{\infty} \sum_{j=0}^{\infty} \psi_j B^{-j} \psi_{j+h} B^{j+h} \\ &= \sigma^2 \sum_{j=0}^{\infty} \psi_j B^{-j} \sum_{h=-\infty}^{\infty} \psi_{j+h} B^{j+h} \\ &= \sigma^2 \sum_{j=0}^{\infty} \psi_j B^{-j} \sum_{h=0}^{\infty} \psi_h B^h \\ &= \sigma^2 \psi(B)\psi(B^{-1}). \end{aligned}$$

On a donc montré que $\gamma(B)$ est identique à la *fonction génératrice des autocovariances*, ce qui établit la proposition. \square

4.6. Les processus ARIMA

Définition 4.13. S'il existe deux entiers positifs d et D tels que $d + D > 0$ et $(w_t = \nabla^d \nabla_s^D z_t; t \in \mathbb{Z})$ est un processus ARMA stationnaire, alors $(z_t; t \in \mathbb{Z})$ est appelé *processus ARIMA* ("integrated autoregressive moving average").

Remarquons que les processus ARIMA ne sont pas des p.s.s. ni même des p.a.s. Si l'on restreint l'ensemble des instants à l'ensemble des naturels \mathbb{N} , par exemple, en posant $z_t = 0$ pour $t < 0$, on obtient un p.a.s. qui n'est pas stationnaire. En guise d'illustration, supposons $d = 1, D = 0$ et $(w_t; t \in \mathbb{Z})$ qui est un b.b.c.s. Si l'on définit $z_t = z_{t-1} + w_t$, on a

$$\begin{aligned} E(z_t) &= 0 \text{ pour tout } t \\ \text{var}(z_0) &= \text{var}(w_0) = \sigma^2, \quad \text{mettons,} \\ \text{var}(z_t) &= (t + 1)\sigma^2. \end{aligned}$$

Charles et Roy (1980) ont développé une théorie dans laquelle la signification des opérateurs est précisée.

Chapitre 5

IDENTIFICATION, ESTIMATION ET DIAGNOSTICS DANS LES MODÈLES DE SÉRIES CHRONOLOGIQUES

Dans ce chapitre on étudie la modélisation des séries chronologiques au moyen des modèles ARMA et ARIMA. D'autres catégories de modèles seront envisagées au chapitre 7. L'approche est essentiellement celle de Box et Jenkins (1976), complétée par Anderson (1977), Roy (1982), Akaike (1974), Ljung et Box (1978). Pour un exposé plus détaillé, voir par exemple Gouriéroux et Monfort (1983) ou Roy (1980).

La méthode se décompose en 7 étapes (Anderson, 1977) :

- 1) familiarisation avec les données,
- 2) analyse préliminaire,
- 3) identification du modèle,
- 4) estimation des paramètres,
- 5) tests d'adéquation du modèle,
- 6) prévision,
- 7) interprétation des résultats.

5.1. Familiarisation avec les données

Le statisticien commence par s'informer sur le domaine dont relèvent les données qu'il a à traiter. Il prend connaissance des théories existantes, des méthodes habituellement utilisées, des objectifs poursuivis par le client qui le consulte. Il s'enquiert de la manière dont les données ont été obtenues, notamment de la précision des mesures, du manque d'homogénéité dans le temps de ces mesures, des événements qui ont pu affecter la grandeur étudiée, de la période des périodicités inhérentes au phénomène.

Il examine ensuite les données et note l'allure générale, l'existence d'une tendance, de composantes périodiques ou quasi périodiques, de l'importance des fluctuations, de la présence de comportements atypiques et de valeurs aberrantes. A cette fin, il est indispensable d'avoir des graphiques des z_t en fonction du temps mais les graphiques relatifs aux différences seront souvent utiles. Les schémas suivants montrent des comportements étranges où il ne faudra pas s'attendre rapidement à des résultats:

Fig. 5.1 Production des citrons en Israël (Raveh, 1980)



Fig. 5.2 Prix d'un produit pétrolier

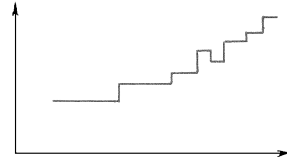


Fig. 5.3 Ventes d'un magasin à rayons multiples après que les cartes perforées portant les données aient été mélangées

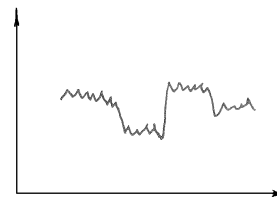
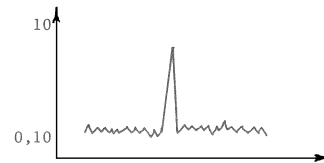


Fig. 5.4 Taux d'intérêt avec une valeur en pourcentage



5.2. Analyse préliminaire

Il s'agit de prendre un certain nombre d'options :

- abandonner des données,
- enquêter au sujet des valeurs aberrantes,
- suppléer les valeurs manquantes,
- transformer les données (logarithmes, inverses, ...),
- choisir de travailler en différences ordinaires (opérateur ∇), différences saisonnières (∇_s) ou les deux différences ($\nabla\nabla_s$) ; à cette fin on peut comparer les variances des séries résultantes et chercher l'opérateur qui minimise la variance; on peut aussi retarder le choix jusqu'à l'étape d'identification,

- extraire une tendance déterministe (ajustement polynomial, par exemple) ou une composante saisonnière déterministe (analyse de Fourier ou régression à l'aide de variables muettes),
- dans le cas multivarié, employer la régression linéaire multiple, y compris en retardant les variables si cela peut se justifier,
- mettre en évidence les analyses d'intervention qui pourront être effectuées.

Ces options ne sont pas définitives. Elles doivent permettre de réaliser l'étape d'identification dans les meilleures conditions possibles.

5.3. Identification du modèle

On peut distinguer trois catégories de méthodes pour la sélection d'un modèle susceptible de représenter une s.c. donnée :

1. Employer une procédure d'élaboration du modèle rigoureuse, telle une procédure de décision statistique multiple. C'est réalisable pour l'ajustement polynomial et pour l'analyse de Fourier (Anderson, 1971, chapitres 3 et 4) à condition que les erreurs aléatoires soient normalement distribuées et indépendantes.
2. Estimer tous les modèles paramétrés appartenant à un ensemble fini et choisir le meilleur au sens d'un certain critère. On peut par exemple chercher le modèle qui minimise la variance résiduelle corrigée pour le biais, ou chercher parmi le modèle qui minimise un critère d'information, tel que AIC (Akaike, 1974)

$$AIC = -2 \ln \hat{L} + 2P$$

où \hat{L} est la fonction de vraisemblance maximisée par rapport aux paramètres et P est le nombre de ces paramètres. Cela implique d'envisager tous les modèles ARMA(p, q) pour $0 \leq p \leq p_{\max}$, $0 \leq q \leq q_{\max}$ et d'estimer les paramètres de ces modèles (Ozaki, 1977).

3. Faire un choix raisonnable de plusieurs modèles en utilisant :
 - le comportement connu de valeurs typiques du modèle théorique; par exemple : la fonction d'autocorrélation d'un modèle ARMA(p, q) stationnaire du second ordre,
 - les estimations de ces valeurs typiques réalisées à partir de la s.c.,
 - la connaissance qu'on a de la distribution échantillonnée des estimateurs.

Nous considérons ici uniquement cette troisième approche.

On définit l'autocorrélation de retard h , ρ_h , par $\rho_h = \gamma_h / \gamma_0$ où γ_h est l'autocovariance de retard h . On l'estime au moyen de

$$\hat{\rho}_h = \frac{\hat{\gamma}_h}{\hat{\gamma}_0}, \quad \text{où} \quad \hat{\gamma}_h = \frac{1}{n} \sum_{t=1}^{n-h} (z_t - \bar{z})(z_{t+h} - \bar{z})$$

où (z_1, \dots, z_n) sont les données et $\bar{z} = \sum z_t/n$. La comparaison entre la fonction d'autocorrélation théorique et la fonction d'autocorrélation estimée peut s'appuyer sur une combinaison de tests statistiques, avec l'inconvénient que ces tests ne sont pas indépendants et donc que le risque de première espèce est inconnu.

Par exemple, on peut tester l'hypothèse nulle H_ℓ que le processus est MA(ℓ) par rapport à l'hypothèse adverse $H_{\ell i} : \rho_i \neq 0$ où $\ell < i \leq \ell_{\max}$, pour plusieurs valeurs de ℓ , où ℓ_{\max} est une fraction de n , pas trop grande. On se base pour cela sur la distribution asymptotique de $\hat{\rho}_i$ sous H_ℓ qui est $N(0, V_\ell/n)$ où $V_\ell = 1 + 2\sum_{j=1}^{\ell} \rho_j^2$ doit être estimé en utilisant les $\hat{\rho}_j$ au lieu des ρ_j . Au niveau de probabilité nominal de 5 %, on rejette H_ℓ en faveur de $H_{\ell i}$ si $|\hat{\rho}_i| > 1.96\sqrt{\hat{V}_\ell/n}$. On peut toutefois s'attendre à rejeter H_ℓ environ dans 5% des cas si H_ℓ est vraie. En prenant successivement $\ell = 0, 1, 2, \dots$, on peut ainsi tenter de détecter l'ordre du processus moyenne mobile. En fait, cette procédure est trop rigide car l'analyste attachera plus d'importance aux retards $1, 2, 3, s-1, s, s+1$ qu'aux autres, où s désigne la période saisonnière.

Dans le raisonnement ci-dessus, on a fait l'hypothèse que le processus est de type MA. Il existe des procédés similaires pour les processus AR (basés sur la fonction d'autocorrélation partielle et plus simples, en fait) et pour les processus ARMA (la méthode du coin de Beguin, Gouriéroux et Monfort, 1980, ou des méthodes similaires, beaucoup plus complexes à mettre en oeuvre que ce qui a été illustré ci-dessus).

Les critiques à l'encontre de cette approche sont nombreuses. Pour ce qui concerne la fonction d'autocorrélation, on constate que le comportement des $\hat{\rho}_\ell$, n'a souvent que peu de ressemblance avec celui des ρ_ℓ . Les causes en sont : le biais des estimateurs, la corrélation qui existe entre eux pour des retards différents et les fluctuations statistiques. De plus les tests sont asymptotiques et le nombre de tests effectués est parfois élevé, ce qui enlève toute valeur au risque de première espèce nominal.

Revenons un instant au choix des degrés de différenciation qui peut également utiliser la fonction d'autocorrélation. Anderson (1979b) préconise de comparer les $\hat{\rho}_\ell$ aux $E(\hat{\gamma}_n/E(\hat{\gamma}_0))$ qui ont un comportement assez typique si le processus générateur est non stationnaire. Roy (1982) propose de calculer les $\hat{\rho}_\ell$ à partir de séries partielles ($z_t : t = 1, \dots, n_j$), $n_1 < n_2 < \dots < n$, et d'examiner si la convergence des graphes est rapide, auquel cas la série peut être considérée comme stationnaire.

5.4. L'estimation des paramètres

Plusieurs méthodes sont utilisées :

1. sans hypothèse de normalité :
 - a. la *méthode des moindres carrés conditionnelle* (MCC, "conditional least squares");
 - b. la *méthode des moindres carrés non conditionnelle* (MCN, "unconditional least squares") préconisée par Box et Jenkins (1976);

2. avec l'hypothèse de normalité :

- a. la *méthode du maximum de vraisemblance conditionnelle* (identique en fait à MCC);
- b. la *méthode du maximum de vraisemblance* (MV, "maximum likelihood") exacte et diverses approximations.

Les propriétés asymptotiques sont identiques dans tous les cas mais Ansley et Newbold (1980) ont mis en évidence les propriétés finies à l'aide de simulations. Leur exemple le plus discriminant est relatif au modèle $w_t = (1 - \theta B)(1 - \Theta B^{12})e_t$ avec $n = 50$ observations (les mauvais résultats s'améliorent nettement quand $n = 100$). Notons $\theta(B) = (1 - \theta B)(1 - \Theta B^{12})$ et C le cercle de rayon 1. Le tableau suivant rassemble les résultats les plus marquants.

Méthode	Caractéristiques	Exemples
MCC	biais si les zéros de $\theta(B)$ sont proches de C	$\theta = 0.85$; $\Theta = 0.85$ $E(\hat{\theta}) = 0,79$; $E(\hat{\Theta}) = 0,58$
MCN	biais si les zéros de $\theta(B)$ sont éloignés de C risque que les zéros soient à l'extérieur de C	$\theta = 0.40$; $\Theta = 0.40$ $E(\hat{\theta}) = 0,43$; $E(\hat{\Theta}) = 0,84$
MV	risque que les zéros soient sur C	

Dans tous les cas, trois types de problèmes numériques se présentent :

1. évaluer la somme des carrés (très facile avec les MCC) ou la fonction de vraisemblance (chapitre 6);
2. minimiser ou maximiser, selon le cas, une fonction de P variables, où P est le nombre de paramètres du modèle, au moyen d'un algorithme itératif assorti d'un critère de convergence;
3. évaluer les dérivées partielles de cette fonction, soit dans le cadre de l'algorithme discuté au point 2, afin d'accélérer la promenade dans l'espace des paramètres, soit après convergence de cet algorithme, dans le but de déterminer une zone de confiance pour les paramètres.

5.5. Tests d'adéquation du modèle

Les tests d'adéquation du modèle sont de plusieurs types :

1. Tests numériques :
 - la convergence de l'algorithme d'optimisation est-elle réalisée?
 - n'y a-t-il pas d'indication qu'un extremum local mais non global ait été atteint?
 - la méthode MCN a-t-elle été correctement exécutée?

2. Tests sur le modèle :

- le modèle est-il stationnaire (après différentiation)?
- le modèle est-il inversible ou situé à la frontière d'inversibilité?
- dans le cas de la méthode MCN , les modules des zéros des polynômes ne sont-ils pas trop proches de 1 ce qui invaliderait la méthode?
- n'y a-t-il pas, exactement ou approximativement, simplification de facteurs entre les polynômes AR et MA ?

3. Tests sur les estimations:

- la distribution asymptotique des estimateurs suggère-t-elle l'emploi d'un modèle simplifié?
- si le modèle est une tentative pour inclure un paramètre supplémentaire, est-ce justifié?

4. Tests sur les résidus:

- la moyenne des résidus est-elle proche de 0 ?
- n'y a-t-il pas un ou plusieurs résidus excessifs (positifs ou négatifs) qui indiqueraient la présence de valeurs aberrantes?
- le graphe des résidus en fonction du temps semble-t-il approximativement compatible avec une suite de variables aléatoires non corrélées?
- les fonctions d'autocorrélation et d'autocorrélation partielle n'indiquent-elles pas la nécessité de compléter le modèle en ajoutant des termes dans les polynômes?
- le test global (Ljung et Box, 1978) sur les autocorrélations résiduelles suggère-t-il de rejeter le modèle?
- l'hypothèse de normalité du processus n'est-elle pas mise en doute?
- le modèle est-il valable de façon homogène au cours du temps?

Au sujet du test Q' de Ljung et Box, qu'il ne faut pas confondre avec le test Q de Box et Pierce (1970) (à proscrire sauf pour des séries très longues), il ne faut pas mésestimer l'importance du nombre L d'autocorrélations utilisées dans la définition

$$Q' = n(n+2) \sum_{\ell=1}^L \frac{\hat{\rho}_{\ell}^2}{n-\ell}.$$

En première approximation, prendre L proche de $n/5$ semble bon si l'on se base sur les simulations produites par Ljung et Box (1978).

5.6. La prévision

Le calcul des prévisions ne pose pas de problème s'il n'y a pas eu de transformation préalable. Notons $(z_t; t \in \mathbb{Z})$ le processus qui nous intéresse et $(w_t; t \in \mathbb{Z})$ le processus transformé au moyen de $z_t = g(w_t)$. Supposons que le p.a. $(w_t; t \in \mathbb{Z})$ soit

gaussien et entièrement spécifié, donc ne dépende pas de paramètres. Supposons de même que la fonction $g(\cdot)$ soit entièrement spécifiée et monotone. Considérons une s.c. générée à partir du p.a. $(z_t; t \in \mathbb{Z})$ que l'on connaît pour $t \leq n$ et que l'on doit prévoir pour $t = n+1, \dots, n+\ell$. Toutes les distributions sont donc conditionnelles à $(z_t; t \leq n)$. Notons $\hat{w}_n(i)$ la prévision d'horizon i faite en n , donc pour prévoir w_{n+i} , et déterminons $\hat{z}_n(i) = g(\hat{w}_n(i))$. Ce qui nous intéresse, c'est la distribution liée de $(z_{n+1}, \dots, z_{n+\ell})$ ou d'une fonction de ces variables, par exemple la valeur cumulée $z_n[\ell] = z_{n+1} + \dots + z_{n+\ell}$. Par hypothèse, la distribution liée de $(w_{n+1}, \dots, w_{n+\ell})$ est normale et le vecteur des moyennes est $(\hat{w}_n(1), \dots, \hat{w}_n(\ell))$. Notons $\mathbf{\Gamma}_\ell$ la matrice de covariance de ces ℓ variables. Par conséquent $\hat{w}_n[\ell] = \hat{w}_n(1) + \dots + \hat{w}_n(\ell)$ est la moyenne de $w_n[\ell] = w_{n+1} + \dots + w_{n+\ell}$ et la variance de $\hat{w}_n[\ell]$ peut être évaluée par la formule $\text{var}(\hat{w}_n[\ell]) = \mathbf{1}'\mathbf{\Gamma}_\ell\mathbf{1}$ où $\mathbf{1} = (1, \dots, 1)^\top$. La distribution liée de $(z_{n+1}, \dots, z_{n+\ell})$ doit alors être déterminée par un procédé numérique. La seule affirmation simple est la suivante. Si l'on note u_α le quantile d'ordre α de la distribution normale centrée réduite et σ_{n+i} l'écart type de z_{n+i} , alors $g(\hat{z}_n(i) + u_\alpha\sigma_{n+i})$ est le quantile d'ordre α de la distribution de z_{n+i} . Le fait que le processus des w_t et la fonction $g(\cdot)$ soient paramétrés et que ces paramètres doivent être estimés complique encore les perspectives.

5.7. Interprétation du modèle

L'interprétation d'un modèle ARMA n'est pas toujours simple:

- interprétation des zéros complexes du polynôme AR donnant lieu à des pseudo-cycles dont la période est de la forme $2\pi/|\omega|$ où ω est l'argument (en radians) des deux zéros imaginaires conjugués, écrits sous la forme $\rho e^{i\omega}$ et $\rho e^{-i\omega}$;
- interprétation de la présence de quelques termes des polynômes en relation avec l'existence de corrélations avec retard.

Cependant, des modèles tels que ceux présentés aux paragraphes 3.5 et 4.5 sont nettement plus riches. On les appellera des modèles interprétables. Sous certaines conditions, ils peuvent être mis sous la forme d'un modèle ARIMA. Ils ne constituent pas la seule voie possible car on peut encore citer les décompositions saisonnières (Hillmer, S.C. and Tiao, G.C., 1982), la décomposition d'une grandeur connue comme somme de plusieurs grandeurs (Granger et Newbold, 1977), la considération d'un modèle univarié comme une équation finale d'un modèle économétrique (Zellner et Palm, 1974). Cependant, dans beaucoup de cas, l'approche de Box et Jenkins peut être utilisée pour cette classe plus large de modèles, en procédant comme suit:

1. identifier, estimer et valider un modèle ARIMA ;
2. rechercher un modèle interprétable parmi ceux envisagés ci-dessus dont la représentation sous forme ARIMA est proche de celle trouvée au point 1 ;

3. estimer les paramètres du modèle interprétable, éventuellement en profitant de la représentation ARIMA, compte tenu des contraintes sur les paramètres qui sont induites par le modèle interprétable.

Cette approche a été suivie par Mélard et Rouland (1986) sur un certain nombre de modèles considérés au paragraphe 3.5.

Chapitre 6

L'ESTIMATION DES PARAMÈTRES DE MODÈLES ARMA

Ce chapitre est consacré en entier aux algorithmes d'estimation des paramètres de modèles ARMA, constants ou évolutifs, univariés ou multivariés. L'estimation est réalisée à l'aide d'un algorithme d'optimisation. Nous n'aborderons pas cet aspect. En revanche, on examine l'évaluation de la fonction de vraisemblance (f.v.) après avoir traité – sur un exemple – du calcul de la somme des carrés conditionnelle (s.c.c.) et de la somme des carrés non conditionnelle (s.c.n.).

La section 6.2 est basée sur Mélard (1982c) qui présente en plus les rapprochements qu'on peut établir entre la méthode de Ljung et Box (1979) et les méthodes proposées antérieurement. La section 6.3 est basée sur Ansley (1979) et Mélard (1982a). La section 6.4 commence par un développement du filtre de Kalman (Kalman, 1960) qui s'inspire de Hannan (1970) (voir aussi Jazwinski, 1970, Anderson et Moore, 1979, Åström, 1970 et Goodwin et Payne, 1977). Pour l'application à l'évaluation de la f.v. par ce procédé, on peut citer Harvey et Phillips (1979) et Ansley et Kohn (1983). L'algorithme rapide qui fait l'objet de la section 6.5 repose sur les travaux de Morf, Sidhu et Kailath (1974), Lindquist (1974), Rissanen (1973), Pearlman (1980) et Mélard (1984).

6.1. Les méthodes basées sur la somme des carrés

Ces méthodes d'évaluation de la somme des carrés sont très simples, bien connues et leur intérêt s'est réduit. Nous allons donc nous contenter de la présenter sur un exemple, celui du processus univarié MA(1).

$$z_t = e_t - \theta e_{t-1},$$

où le coefficient θ est constant et les e_t sont les innovations, de variance σ^2 .

La méthode MCC basée sur la somme des carrés conditionnelle (s.c.c.) consiste à calculer

$$S(\theta) = \sum_{t=1}^n e_t^2(\theta),$$

où $e_t(\theta)$ n'est autre que e_t calculé par récurrence, avec une valeur initiale $e_0(\theta) = 0$, et considérée comme une fonction de θ . On a donc

$$\begin{aligned} e_1(\theta) &= z_1, \\ e_2(\theta) &= z_2 + \theta e_1(\theta) = z_2 + \theta z_1, \\ e_3(\theta) &= z_3 + \theta e_2(\theta) = z_3 + \theta z_2 + \theta^2 z_1, \\ &\dots \\ e_t(\theta) &= z_t + \theta e_{t-1}(\theta) = \sum_{j=0}^{t-1} \theta^j z_{t-j}. \end{aligned} \tag{6.1}$$

Notons que $e_t(\theta)$ n'est pas une fonction linéaire de θ . C'est d'ailleurs un des rares cas où il est possible d'écrire l'expression explicite de $S(\theta)$

$$S(\theta) = \sum_{t=1}^n \left(\sum_{j=0}^{t-1} \theta^j z_{t-j} \right)^2.$$

La fonction de vraisemblance conditionnelle (f.v.c.), étant donnée $e_0 = 0$, est liée à $S(\theta)$, sous l'hypothèse complémentaire que $\mathbf{z} = (z_1, \dots, z_n)^\top$ a une distribution normale. En effet, cette f.v.c. peut s'écrire

$$L(\theta, \sigma^2; \mathbf{z} | e_0 = 0) = f(\mathbf{z}; \theta, \sigma^2 | e_0 = 0), \tag{6.2}$$

où $f(\mathbf{v}; \boldsymbol{\beta} | \mathbf{u} = \mathbf{u}_0)$ représente, de façon générale, la fonction de densité de \mathbf{v} conditionnellement à $\mathbf{u} = \mathbf{u}_0$, dépendant des paramètres spécifiés par $\boldsymbol{\beta}$. On effectue le changement de variable spécifié par (6.1), qui s'écrit $\mathbf{e} = \mathbf{T}\mathbf{z}$ ou $\mathbf{z} = \mathbf{T}^{-1}\mathbf{e}$, avec \mathbf{T} qui est une matrice $n \times n$ triangulaire inférieure:

$$\mathbf{T} = \begin{bmatrix} 1 & & & \\ \theta & 1 & & 0 \\ \theta^2 & \theta & 1 & \\ \dots & & & \ddots \\ \theta^{n-1} & \dots & & \theta & 1 \end{bmatrix}.$$

Notons que le jacobien de la transformation vaut $J(\theta) = |\det \mathbf{T}| = 1$. Les e_t ont une distribution liée normale et sont non corrélées. Ce sont donc des v.a. indépendantes, normalement distribuées, de moyenne zero et de variance σ^2 . D'où (6.2) devient

$$\begin{aligned} L(\theta, \sigma^2; \mathbf{z}|e_0 = 0) &= J(\theta) \prod_{t=1}^n \left\{ (2\pi)^{-1/2} \sigma^{-1} \exp\left(-\frac{e_t^2}{2\sigma^2}\right) \right\} \\ &= (2\pi)^{-n/2} \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{t=1}^n e_t^2\right\}, \end{aligned}$$

où l'on écrit e_t pour $e_t(\theta)$. Le logarithme de la f.v.c. s'écrit donc, à une constante additive près,

$$\ell(\theta, \sigma^2; \mathbf{z}|e_0 = 0) = -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^n e_t^2.$$

L'équation de vraisemblance correspondant à σ^2 est

$$\frac{\partial \ell}{\partial \sigma^2} = 0 \quad \text{ou} \quad -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{t=1}^n e_t^2 = 0$$

dont la solution est

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^n e_t^2 = \frac{1}{n} S(\theta).$$

Par conséquent, on est ramené à maximiser le *logarithme de la fonction de vraisemblance concentrée* :

$$\ell(\theta, \hat{\sigma}^2; \mathbf{z}|e_0 = 0) = -\frac{n}{2} \log \hat{\sigma}^2 - \frac{n}{2},$$

ce qui revient à maximiser $-\log \hat{\sigma}^2$ ou à minimiser la s.c.c. $S(\theta)$.

La méthode MCN basée sur la somme des carrés non conditionnelle (s.c.n.) consiste, pour le modèle MA(1) considéré ici, à supprimer la condition portant sur e_0 , en reculant d'un rang dans le passé. On remplace donc la première relation de (6.1) par $e_1(\theta) = z_1 + \theta \hat{e}_0(\theta)$ et on ajoute une relation $e_0(\theta) = \hat{z}_0$. On évalue \hat{z}_0 comme prévision à reculons, obtenue à partir du modèle en temps inversé

$$z_t = e'_t - \theta e'_{t+1},$$

où e'_t est l'innovation en temps inversé au temps t . Dès lors,

$$\hat{z}_0 = -\theta e'_1 = -\theta \sum_{j=0}^{n-1} \theta^j z_{1+j} = -\sum_{t=1}^n \theta^t z_t. \quad (6.3)$$

La méthode basée sur la s.c.n. n'est pas équivalente à la méthode du m.v., pour deux raisons essentiellement:

1. \hat{z}_0 n'est pas l'espérance mathématique de z_0 , conditionnellement à \mathbf{z} , dont on peut vérifier qu'elle vaut

$$-\sum_{t=1}^n \theta^{-t} \frac{\theta^{2t} - \theta^{2n+2}}{1 - \theta^{2n+2}} z_t$$

ce qui converge vers (6.3) quand $n \rightarrow \infty$, en supposant $|\theta| < 1$;

2. $e_t(\theta)$ dépend de tous les z_t , par l'intermédiaire de \hat{z}_0 ; par conséquent, la matrice jacobienne n'est pas triangulaire; en fait, le jacobien vaut

$$\left(\frac{1 - \theta^2}{1 - \theta^{2n+2}} \right)^{1/2},$$

ce qui ne tend pas vers 1 quand $n \rightarrow \infty$.

6.2. La méthode de Ljung et Box

La méthode de Ljung et Box (1979) est l'aboutissement d'une succession de méthodes parues depuis Newbold (1974) et qui visent à évaluer la f.v. exacte dans le cadre de la méthode MV. Appliquons-la directement à un processus ARMA(p, q). Pour calculer les e_t à partir des z_t au moyen de la relation

$$z_t - \phi_1 z_{t-1} - \dots - \phi_p z_{t-p} = e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}, \quad (6.4)$$

il est nécessaire de disposer de valeurs initiales, $\mathbf{v}_0 = (z_0^*, \dots, z_{1-p}^*, e_0^*, \dots, e_{1-q}^*)^\top$. En écrivant (6.4) pour $t = 1, \dots, n$, on obtient la relation matricielle

$$\mathbf{A}_1 \mathbf{z} = \mathbf{A}_2^{-1} \mathbf{e} + \mathbf{B} \mathbf{v}_0 \quad (6.5)$$

où $\mathbf{A}_1, \mathbf{A}_2^{-1}$ sont des matrices triangulaires inférieures, comportant des 1 sur la diagonale principale (donc ce sont des matrices inversibles) et \mathbf{B} est une matrice $n \times h$, avec $h = p + q$, dont seules les $m = \max(p, q)$ premières lignes sont non nulles car les valeurs initiales n'interviennent directement dans l'expression de z_t que pour $t \leq m$. Notons $\sigma^2 \boldsymbol{\Sigma}$ la matrice de covariance de \mathbf{v}_0 . Puisque les e_t sont des v.a. non corrélées avec le passé du p.a., on a $\mathbf{e} \perp \mathbf{v}_0$ et donc

$$\|\mathbf{A}_1 \mathbf{z}\|^2 = \sigma^2 (\mathbf{A}_2^{-1} \mathbf{A}_2^{-1\top} + \mathbf{B} \boldsymbol{\Sigma} \mathbf{B}^\top).$$

Or $\|\mathbf{A}_1 \mathbf{z}\|^2 = \sigma^2 \mathbf{A}_1 \boldsymbol{\Lambda} \mathbf{A}_1^\top$ si l'on note $\sigma^2 \boldsymbol{\Lambda}$ la matrice de covariance des z_t ($t = 1, \dots, n$). D'où

$$\boldsymbol{\Lambda} = \mathbf{A}_1^{-1} (\mathbf{A}_2^{-1} \mathbf{A}_2^{-1\top} + \mathbf{B} \boldsymbol{\Sigma} \mathbf{B}^\top) \mathbf{A}_1^{-1\top}.$$

Notons $\boldsymbol{\beta} = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)^\top$ le vecteur des paramètres. Sous l'hypothèse que \mathbf{z} a une distribution normale, de vecteur moyenne $\mathbf{0}$ et de matrice de covariance $\sigma^2 \boldsymbol{\Lambda}$, la f.v. s'écrit

$$L(\boldsymbol{\beta}, \sigma^2; \mathbf{z}) = (2\pi)^{-n/2} \sigma^{-n} (\det \boldsymbol{\Lambda})^{-1/2} \exp \left(-\frac{1}{2\sigma^2} \mathbf{z}^\top \boldsymbol{\Lambda}^{-1} \mathbf{z} \right). \quad (6.6)$$

On peut déterminer $\boldsymbol{\Lambda}^{-1}$ à partir de

$$\boldsymbol{\Lambda}^{-1} = \mathbf{A}_1^\top \mathbf{A}_2^\top (\mathbf{I}_n + \mathbf{A}_2 \mathbf{B} \boldsymbol{\Sigma} \mathbf{B}^\top \mathbf{A}_2^\top)^{-1} \mathbf{A}_2 \mathbf{A}_1,$$

d'où

$$\mathbf{z}^\top \mathbf{A}^{-1} \mathbf{z} = \tilde{\mathbf{e}}^\top (\mathbf{I}_n + \mathbf{A}_2 \mathbf{B} \Sigma \mathbf{B}^\top \mathbf{A}_2^\top)^{-1} \tilde{\mathbf{e}}$$

ou $\tilde{\mathbf{e}} = \mathbf{A}_2 \mathbf{A}_1 \mathbf{z}$ est le vecteur des e_t que l'on obtient en remplaçant dans (6.5), \mathbf{v}_0 par $\mathbf{0}$, c'est-à-dire en appliquant la méthode conditionnelle de la section 6.1. Le but est d'éviter l'inversion d'une matrice $n \times n$. On part de la remarque suivante : le produit $\mathbf{A}_2 \mathbf{B}$ peut être remplacé par le produit $\mathbf{A}_2^* \mathbf{B}^*$ où \mathbf{B}^* est la matrice $m \times h$ des m seules lignes non nulles de \mathbf{B} et \mathbf{A}_2^* est la matrice $n \times m$ constituée des m premières colonnes de \mathbf{A}_2 . Cette substitution conduit à

$$\mathbf{z}^\top \mathbf{A}^{-1} \mathbf{z} = \tilde{\mathbf{e}}^\top (\mathbf{I}_n + \mathbf{A}_2^* \mathbf{B}^* \Sigma \mathbf{B}^{*\top} \mathbf{A}_2^{*\top})^{-1} \tilde{\mathbf{e}}.$$

On a déjà utilisé précédemment une règle d'inversion pour matrice partitionnée (4.14). Avec les mêmes notations mais en permutant les indices 1 et 2, on arrive pour \mathbf{A}^{22} à l'expression

$$\mathbf{A}_{22}^{-1} + \mathbf{A}_{22}^{-1} \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \quad \text{avec} \quad \mathbf{A}^{11} = (\mathbf{A}_{11} - \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{A}_{21})^{-1}.$$

En égalant les deux expressions de \mathbf{A}^{22} , on obtient

$$(\mathbf{A}_{22} - \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{A}_{12})^{-1} = \mathbf{A}_{22}^{-1} + \mathbf{A}_{22}^{-1} \mathbf{A}_{21} (\mathbf{A}_{11} - \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{A}_{21})^{-1} \mathbf{A}_{12} \mathbf{A}_{22}^{-1}.$$

On applique ici cette identité avec

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} = \begin{bmatrix} -(\mathbf{B}^* \Sigma \mathbf{B}^{*\top})^{-1} \mathbf{A}_2^{*\top} \\ \mathbf{A}_2^* & \mathbf{I}_n \end{bmatrix}$$

pour obtenir

$$\mathbf{z}^\top \mathbf{A}^{-1} \mathbf{z} = \tilde{\mathbf{e}}^\top [\mathbf{I}_n - \mathbf{A}_2^* \{(\mathbf{B}^* \Sigma \mathbf{B}^{*\top})^{-1} + \mathbf{A}_2^{*\top} \mathbf{A}_2^*\}^{-1} \mathbf{A}_2^{*\top}] \tilde{\mathbf{e}},$$

ce qui ne nécessite plus que deux inversions de matrices $m \times m$. Il n'est pas étonnant que les éléments de \mathbf{A}_2^* puissent s'obtenir assez aisément puisque la matrice \mathbf{A}_2^{-1} qu'on avait au départ est une matrice triangulaire et de bande. De plus (4.14) implique que

$$\begin{aligned} \det(\mathbf{A}) &= \det(\mathbf{A}_{11}) \det(\mathbf{A}_{22} - \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{A}_{12}) \\ &= \det(\mathbf{A}_{22}) \det(\mathbf{A}_{11} - \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{A}_{21}), \end{aligned}$$

d'où, avec les mêmes notations que ci-dessus,

$$\det(\mathbf{I}_n + \mathbf{A}_2^* \mathbf{B}^* \Sigma \mathbf{B}^{*\top} \mathbf{A}_2^{*\top}) = \det\{(\mathbf{B}^* \Sigma \mathbf{B}^{*\top}) \det(\mathbf{B}^* \Sigma \mathbf{B}^{*\top})^{-1} + \mathbf{A}_2^{*\top} \mathbf{A}_2^*\}.$$

Le développement présenté ici est basé sur un article (Mélard, 1982c) qui présente notamment la méthode de Ljung et Box sous une forme plus simple que dans l'article original et dans lequel les liens avec un certain nombre d'autres méthodes sont explicites (voir aussi Kestemont, 1983). La méthode de Ljung et

Box a été généralisée aux processus MA multivariés par Hillmer et Tiao (1979). Mis sous la forme présente, l'algorithme peut être généralisé aux processus ARMA évolutifs. Les différences essentielles sont les suivantes :

- \mathbf{I}_n doit être remplacé par une matrice diagonale dont les éléments sont inversement proportionnels aux variances des innovations;
- le calcul de \mathbf{A}_2^* , non détaillé ici, devient plus compliqué.

6.3. La méthode de factorisation de Cholesky

La méthode de Ansley (1979) est très simple dans son principe. Elle est basée sur la factorisation de Cholesky de $\mathbf{\Lambda}$ (Voir Section 6.2), qui est définie positive : $\mathbf{\Lambda} = \mathbf{P}\mathbf{P}^\top$, où \mathbf{P} est une matrice triangulaire inférieure dont les éléments diagonaux p_{ii} sont strictement positifs. Ansley a évidemment présenté la méthode dans le cas de processus ARMA constants mais elle est conceptuellement presque inchangée pour les processus ARMA évolutifs, en supposant que des hypothèses ont permis de déterminer $\mathbf{\Lambda}$ (voir Section 3.6).

Considérons la décomposition de Wold-Cramér du p.p.i. $(z_t; t \geq 1)$. Notons l'innovation en t sous la forme $h_t \tilde{e}_t$, où $\text{var}(\tilde{e}_t) = \sigma^2$; $h_t \sigma$ est donc l'écart type de cette innovation. La décomposition de z_t s'écrit

$$z_t = h_t \tilde{e}_t + \sum_{j=1}^{t-1} \varphi_{jt} (h_{t-j} \tilde{e}_{t-j}).$$

Afin de mettre cette relation sous forme matricielle, notons \mathbf{H} la matrice diagonale $n \times n$ comportant les h_t dans l'ordre croissant, $\tilde{\mathbf{e}} = (\tilde{e}_1, \dots, \tilde{e}_n)^\top$ et $\boldsymbol{\varphi}$ la matrice triangulaire inférieure comportant les φ_{jt} . Plus précisément, la t -ième ligne de $\boldsymbol{\varphi}$ s'écrit:

$$(\varphi_{-1,t}, \dots, \varphi_{1t}, 1, 0, \dots, 0).$$

Dès lors, il vient

$$\mathbf{z} = \boldsymbol{\varphi} \mathbf{H} \tilde{\mathbf{e}} \quad \text{et} \quad \tilde{\mathbf{e}} = \mathbf{H}^{-1} \boldsymbol{\varphi}^{-1} \mathbf{z}.$$

Puisque les \tilde{e}_t sont des v.a. non corrélées, de variance σ^2 , on a successivement

$$\begin{aligned} \mathbf{\Lambda} &= \sigma^{-2} \|\mathbf{z}\|^2 = \boldsymbol{\varphi} \mathbf{H} \mathbf{H}^\top \boldsymbol{\varphi}^\top, \\ \mathbf{\Lambda}^{-1} &= \boldsymbol{\varphi}^{-1\top} \mathbf{H}^{-1\top} \mathbf{H}^{-1} \boldsymbol{\varphi}^{-1}, \\ \mathbf{z}^\top \mathbf{\Lambda}^{-1} \mathbf{z} &= \mathbf{z}^\top \boldsymbol{\varphi}^{-1\top} \mathbf{H}^{-1\top} \mathbf{H}^{-1} \boldsymbol{\varphi}^{-1} \mathbf{z} = \tilde{\mathbf{e}}^\top \tilde{\mathbf{e}} = \sum_{t=1}^n \tilde{e}_t^2, \end{aligned}$$

et

$$\det(\mathbf{\Lambda}) = (\det \boldsymbol{\varphi})^2 (\det \mathbf{H})^2 = \prod_{t=1}^n h_t^2.$$

Il suffit donc de calculer les innovations normalisées \tilde{e}_t et les h_t . Par substitution dans (6.6), on obtient :

$$L(\boldsymbol{\beta}, \sigma^2; \mathbf{z}) = (2\pi)^{-n/2} \sigma^{-n} \left(\prod_{t=1}^n h_t^2 \right)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{t=1}^n \tilde{e}_t^2 \right\}.$$

Il s'ensuit que

$$\ell(\boldsymbol{\beta}, \sigma^2; \mathbf{z}) = -\frac{n}{2} \log \sigma^2 - \frac{1}{2} \log \left(\prod_{t=1}^n h_t^2 \right) - \frac{1}{2\sigma^2} \sum_{t=1}^n \tilde{e}_t^2.$$

L'équation de vraisemblance relative à σ^2 donne comme solution $\hat{\sigma}^2 = \sum \tilde{e}_t^2 / n$, d'où le logarithme de la f.v. concentrée est

$$\begin{aligned} \ell(\boldsymbol{\beta}, \sigma^2; \mathbf{z}) &= -\frac{n}{2} \log \left(\frac{1}{n} \sum_{t=1}^n \tilde{e}_t^2 \right) - \frac{1}{2} \log \left(\prod_{t=1}^n h_t^2 \right) - \frac{n}{2} \\ &= -\frac{n}{2} \log \left[\left(\frac{1}{n} \sum_{t=1}^n \tilde{e}_t^2 \right) \left(\prod_{t=1}^n h_t^2 \right)^{1/n} \right] - \frac{n}{2}. \end{aligned} \quad (6.7)$$

Il suffit donc de minimiser $\sum_{t=1}^n \tilde{e}_t^2$ où $\tilde{e}_t = e_t (\prod_{i=1}^t h_i)^{1/n}$.

En dépit de la simplicité apparente, rien n'est gagné en termes d'efficacité de l'algorithme car le calcul de \mathbf{H} et de $\boldsymbol{\phi}$ est relativement laborieux. C'est ici qu'intervient l'idée fondamentale de Ansley : s'arranger pour que $\boldsymbol{\phi}$ soit une matrice de bande dont seulement un nombre restreint r de parallèles à la diagonale principale soient non nulles. Il suffit pour cela que le processus soit $(r-1)$ -dépendant, où $r \leq n$. Partant de $(z_t; t \geq 1)$, on peut obtenir un processus $(r-1)$ -dépendant de bien des manières. Par exemple, on peut poser

$$x_t = \begin{cases} z_t, & t = 1, \dots, p \\ z_t - \phi_{1t} z_{t-1} - \dots - \phi_{pt} z_{t-p}, & t = p+1, \dots, n \end{cases}$$

Le retard maximum dans la décomposition de Wold-Cramér, r , sera de $p-1$ pour x_p et de q pour $x_t, t \geq p+1$, puisque x_t peut se mettre sous la forme MA(q) :

$$x_t = e_t - \theta_{1t} e_{t-1} - \dots - \theta_{qt} e_{t-q}.$$

On a donc $r = \max(p, q-1)$. La transformation de \mathbf{z} à $\mathbf{x} = (x_1, \dots, x_n)^\top$ étant de jacobien 1, le seul problème est donc d'évaluer ce que devient $\boldsymbol{\Lambda} = \|\mathbf{x}\|^2 / \sigma^2$. On se base sur la section 3.6 mais les détails ne seront pas donnés ici (Mélard, 1982a).

6.4. La méthode basée sur le filtre de Kalman

Considérons le modèle linéaire dynamique (m.l.d., Définition 3.28) dont nous rap- pelons ici les hypothèses, pour faciliter les références:

$$(H_1) \quad \mathbf{w}_t = \mathbf{F}_t \mathbf{w}_{t-1} + \mathbf{G}_t \mathbf{e}_t \quad (\text{équation de transition})$$

$$r \times 1 \quad r \times r \quad r \times 1 \quad r \times m \quad m \times 1$$

$$(H_2) \quad \mathbf{z}_t = \mathbf{H}_t \mathbf{w}_t + \mathbf{a}_t \quad (\text{équation d'observation})$$

$$k \times 1 \quad k \times r \quad r \times 1 \quad k \times 1$$

(H3) \mathbf{F}_t , \mathbf{G}_t et \mathbf{H}_t sont des matrices non aléatoires

(H4) \mathbf{z}_t , \mathbf{w}_t , \mathbf{e}_t et \mathbf{a}_t sont des v.a. vectorielles

(H5) $\mathbf{e}_t \perp \mathbf{e}_s, \forall t \neq s$, et $E(\mathbf{e}_t) = 0, \forall t$

(H6) $\|\mathbf{e}_t\|^2 = \mathbf{\Sigma}_t$

(H7) $\mathbf{a}_t \perp \mathbf{a}_s, \forall t \neq s$, et $E(\mathbf{a}_t) = 0, \forall t$

(H8) $\|\mathbf{a}_t\|^2 = \mathbf{S}_t$

(H9) $\mathbf{a}_t \perp \mathbf{e}_s, \forall t, s$

(H10) $E(\mathbf{w}_0) = 0$ et $\|\mathbf{w}_0\| = \mathbf{\Sigma}_0$

(H11) $\mathbf{w}_0 \perp \mathbf{e}_t, \forall t$, et $\mathbf{w}_0 \perp \mathbf{a}_t, \forall t$.

Introduisons les notations suivantes :

$V_t = V_2(\mathbf{z}; t)$ est le sous-espace de Hilbert sous-tendu par

$$\{(\mathbf{z}_s)_i; s \leq t; i = 1, \dots, k\} \quad \text{et} \quad V_0 = \{0\};$$

U_t : le sous-espace de Hilbert sous-tendu par

$$\{(\mathbf{w}_0)_i; i = 1, \dots, k\} \cup \{(\mathbf{e}_s)_i; s \leq t, \quad i = 1, \dots, k\}$$

$$\cup \{(\mathbf{a}_s)_i; s \leq t, \quad i = 1, \dots, k\};$$

$\mathbf{v}_t = P(\mathbf{v}|V_t)$: la projection orthogonale de \mathbf{v} dans V_t ;

$\hat{\mathbf{e}}_t = \mathbf{z}_t - \mathbf{z}_{t|t-1}$: l'innovation de \mathbf{z}_t au temps t ;

$\hat{\mathbf{\Sigma}}_t = \|\hat{\mathbf{e}}_t\|^2$;

I_t : le sous-espace de Hilbert sous-tendu par $\{(\hat{\mathbf{e}}_t)_i; i = 1, \dots, k\}$;

$\mathbf{P}_{t|t-1} = \|\mathbf{w}_t - \mathbf{w}_{t|t-1}\|^2, (t \geq 1)$ et $\mathbf{P}_{0|-1} = \mathbf{\Sigma}_0$.

Lemme 6.1.

1. $V_{t-1} \subset V_t \subset U_t \subset U_{t+1}$,
2. $\mathbf{w}_t \perp \mathbf{e}_t$.

Démonstration.

1. On a (a) $V_{t-1} \subset V_t$ par la Proposition 2.35; (b) $U_t \subset U_{t+1}$ en vertu d'arguments similaires; (c) $V_t \subset U_t$ car (H_1) peut s'écrire sous forme de la décomposition de Wold-Cramér de $(\mathbf{w}_t; t \geq 0)$:

$$\mathbf{w}_t = \sum_{j=0}^{t-1} \left(\prod_{i=0}^{j-1} \mathbf{F}_{t-i} \right) \mathbf{G}_{t-j} \mathbf{e}_{t-j} + \left(\prod_{i=0}^{t-1} \mathbf{F}_{t-i} \right) \mathbf{w}_0, \quad (6.8)$$

d'où les composantes de $\mathbf{z}_t = \mathbf{H}_t \mathbf{w}_t + \mathbf{e}_t$ sont des combinaisons linéaires des v.a. qui sous-tendent U_t . La proposition résulte alors du fait que U_t étant un sous-espace de Hilbert, il est fermé.

2. La deuxième partie de la proposition résulte directement de (H_9) et de (H_{11}) , compte tenu de (6.8). \square

Lemme 6.2. Pour tout $t \geq 1$ et tout $h \geq 0$, on a

- (a) $\mathbf{e}_{t+h} \perp V_{t-1}$,
- (b) $\mathbf{a}_{t+h} \perp V_{t-1}$.

Démonstration. En effet :

$$\begin{aligned} \mathbf{e}_{t+h} \perp \mathbf{w}_0 \quad \text{et} \quad \mathbf{a}_{t+h} \perp \mathbf{w}_0 & \quad \text{par } (H_{11}) \\ \mathbf{e}_{t+h}, \mathbf{a}_{t+h} \perp \mathbf{e}_s, \quad s \leq t-1, & \quad \text{par } (H_5) \text{ et } (H_9), \\ \mathbf{e}_{t+h}, \mathbf{a}_{t+h} \perp \mathbf{a}_s, \quad s \leq t-1, & \quad \text{par } (H_9) \text{ et } (H_7), \end{aligned}$$

donc $\mathbf{e}_{t+h}, \mathbf{a}_{t+h} \perp U_{t-1}$. Du Lemme 6.1, on déduit que $\mathbf{e}_{t+h}, \mathbf{a}_{t+h} \perp V_{t-1}$ puisque $V_{t-1} \subset U_{t-1}$. \square

Lemme 6.3. $(\mathbf{w}_{t|t-1})_i \in V_{t-1} (i = 1, \dots, k)$ et $\mathbf{w}_t - \mathbf{w}_{t|t-1} \perp V_{t-1}$.

Démonstration. En effet, $\mathbf{w}_{t|t-1} = P(\mathbf{w}_t | V_{t-1})$ a ces propriétés grâce au théorème de projection (Proposition 2.17). \square

Lemme 6.4. $\hat{\mathbf{e}}_t = \mathbf{z}_t - \mathbf{H}_t \mathbf{w}_{t|t-1} = \mathbf{H}_t (\mathbf{w}_t - \mathbf{w}_{t|t-1}) + \mathbf{a}_t$ et $\hat{\mathbf{e}} \in V_t$.

Démonstration. En effet

$$\mathbf{z}_{t|t-1} = P(\mathbf{z}_t | V_{t-1}) = P(\mathbf{H}_t \mathbf{w}_t + \mathbf{e}_t | V_{t-1}) = \mathbf{H}_t \mathbf{w}_{t|t-1}$$

grâce au Lemme 6.2. \square

Lemme 6.5. $\hat{\Sigma}_t = \mathbf{H}_t \mathbf{P}_{t|t-1} \mathbf{H}_t^\top + \mathbf{S}_t$.

Démonstration.

$$\begin{aligned}\hat{\Sigma}_t &= \|\hat{\mathbf{e}}_t\|^2 = \|\mathbf{H}_t(\mathbf{w}_t - \mathbf{w}_{t|t-1}) + \mathbf{a}_t\|^2 \\ &= \mathbf{H}_t\|\mathbf{w}_t - \mathbf{w}_{t|t-1}\|^2\mathbf{H}_t^\top + \|\mathbf{e}_t\|^2 + \langle \mathbf{H}_t(\mathbf{w}_t - \mathbf{w}_{t|t-1}), \mathbf{a}_t \rangle + \langle \mathbf{a}_t, \mathbf{H}_t(\mathbf{w}_t - \mathbf{w}_{t|t-1}) \rangle \\ &= \mathbf{H}_t\mathbf{P}_{t|t-1}\mathbf{H}_t^\top + \mathbf{S}_t,\end{aligned}$$

les deux derniers termes disparaissant grâce aux Lemmes 6.1 à 6.3. \square

Proposition 6.6. *Supposons que $\hat{\Sigma}_t = \mathbf{H}_t\mathbf{P}_{t|t-1}\mathbf{H}_t^\top + \mathbf{S}_t$ soit une matrice inversible. Soit la matrice $r \times k$*

$$\mathbf{K}_t = \mathbf{F}_{t+1}\mathbf{P}_{t|t-1}\mathbf{H}_t^\top\hat{\Sigma}_t^{-1} \quad (6.9)$$

dite matrice de gain. Les relations suivantes sont alors vérifiées pour tout $t \geq 1$:

$$(T_1) \quad \mathbf{w}_{t+h|t} = \mathbf{F}_{t+h}\mathbf{w}_{t+h-1|t}, \quad \forall h > 1$$

$$(T_2) \quad \mathbf{w}_{t+1|t} = (\mathbf{F}_{t+1} - \mathbf{K}_t\mathbf{H}_t)\mathbf{w}_{t|t-1} + \mathbf{K}_t\mathbf{z}_t$$

$$(T_3) \quad \mathbf{P}_{t+1|t} = \mathbf{F}_{t+1}\mathbf{P}_{t|t-1}\mathbf{F}_{t+1}^\top + \mathbf{G}_{t+1}\Sigma_{t+1}\mathbf{G}_{t+1}^\top - \mathbf{K}_t\hat{\Sigma}_t\mathbf{K}_t^\top$$

$$(T_4) \quad \mathbf{P}_{1|0} = \mathbf{F}_1\Sigma_0\mathbf{F}_1^\top + \mathbf{G}_1\Sigma_1\mathbf{G}_1^\top$$

Démonstration. La démonstration est subdivisée en plusieurs parties:

(a) Démonstration de (T₁) :

$$\mathbf{w}_{t+h|t} = P(\mathbf{w}_{t+h}|V_t) = P(\mathbf{F}_{t+h}\mathbf{w}_{t+h-1} + \mathbf{G}_{t+h}\mathbf{e}_{t+h}|V_t) = \mathbf{F}_{t+h}\mathbf{w}_{t+h-1|t}$$

grâce au Lemme 6.2.

(b) Démonstration de (T₂) : Par la Proposition 2.38 , on a $V_t = V_{t-1} \oplus I_t$, d'où

$$\mathbf{w}_{t+1|t} = P(\mathbf{w}_{t+1}|V_t) = P(\mathbf{w}_{t+1}|V_{t-1}) + P(\mathbf{w}_{t+1}|I_t).$$

Notons $\mathbf{u}_t = P(\mathbf{w}_{t+1}|I_t)$ que la Proposition 2.24 permet d'écrire

$$\mathbf{u}_t = \langle \mathbf{w}_{t+1}, \hat{\mathbf{e}}_t \rangle \hat{\Sigma}_t^{-1} \hat{\mathbf{e}}_t.$$

Or (H₁) et le Lemme 6.4 entraînent que

$$\begin{aligned}\langle \mathbf{w}_{t+1}, \hat{\mathbf{e}}_t \rangle &= \langle \mathbf{F}_{t+1}\mathbf{w}_t + \mathbf{G}_{t+1}\mathbf{e}_{t+1}, \mathbf{H}_t(\mathbf{w}_t - \mathbf{w}_{t|t-1}) + \mathbf{a}_t \rangle \\ &= \langle \mathbf{F}_{t+1}(\mathbf{w}_t - \mathbf{w}_{t|t-1}) + \mathbf{F}_{t+1}\mathbf{w}_{t|t-1} + \mathbf{G}_{t+1}\mathbf{e}_{t+1}, \mathbf{H}_t(\mathbf{w}_t - \mathbf{w}_{t|t-1}) + \mathbf{a}_t \rangle \\ &= \mathbf{F}_{t+1}\mathbf{P}_{t|t-1}\mathbf{H}_t^\top\end{aligned}$$

par application des Lemmes 6.1, 6.2 et 6.3 et de (H₉). On a donc

$$\mathbf{u}_t = \mathbf{F}_{t+1}\mathbf{P}_{t|t-1}\mathbf{H}_t^\top\hat{\Sigma}_t^{-1}\hat{\mathbf{e}}_t = \mathbf{K}_t\hat{\mathbf{e}}_t$$

et, par conséquent,

$$\begin{aligned}\mathbf{w}_{t+1|t} &= \mathbf{w}_{t+1|t-1} + \mathbf{K}_t \widehat{\mathbf{e}}_t \\ &= \mathbf{F}_{t+1} \mathbf{w}_{t|t-1} + \mathbf{K}_t (\mathbf{z}_t - \mathbf{H}_t \mathbf{w}_{t|t-1}),\end{aligned}$$

grâce à (T₁) et au Lemme 6.4 , puis

$$\mathbf{w}_{t+1|t} = (\mathbf{F}_{t+1} - \mathbf{K}_t \mathbf{H}_t) \mathbf{w}_{t|t-1} + \mathbf{K}_t \mathbf{z}_t.$$

(c) Démonstration de (T₃) :

$$\begin{aligned}\mathbf{P}_{t+1|t} &= \|\mathbf{w}_{t+1} - \mathbf{w}_{t+1|t}\|^2 \\ &= \|\mathbf{F}_{t+1}(\mathbf{w}_t - \mathbf{w}_{t|t-1}) + \mathbf{G}_{t+1} \mathbf{e}_{t+1} - \mathbf{K}_t \widehat{\mathbf{e}}_t\|^2 \quad (\text{par } H_1 \text{ et } T_2) \\ &= \mathbf{F}_{t+1} \|\mathbf{w}_t - \mathbf{w}_{t|t-1}\|^2 \mathbf{F}_{t+1}^\top + \mathbf{G}_{t+1} \|\mathbf{e}_{t+1}\|^2 \mathbf{G}_{t+1}^\top \\ &\quad + \langle \mathbf{F}_{t+1}(\mathbf{w}_t - \mathbf{w}_{t|t-1}), \mathbf{G}_{t+1} \mathbf{e}_{t+1} \rangle \\ &\quad + \langle \mathbf{G}_{t+1} \mathbf{e}_{t+1}, \mathbf{F}_{t+1}(\mathbf{w}_t - \mathbf{w}_{t|t-1}) \rangle \\ &\quad - \langle \mathbf{w}_{t+1} - \mathbf{w}_{t+1|t}, \mathbf{K}_t \widehat{\mathbf{e}}_t \rangle - \langle \mathbf{K}_t \widehat{\mathbf{e}}_t, \mathbf{w}_{t+1} - \mathbf{w}_{t+1|t} \rangle - \mathbf{K}_t \|\widehat{\mathbf{e}}_t\|^2 \mathbf{K}_t \\ &= \mathbf{F}_{t+1} \mathbf{P}_{t|t-1} \mathbf{F}_{t+1}^\top + \mathbf{G}_{t+1} \boldsymbol{\Sigma}_{t+1} \mathbf{G}_{t+1}^\top - \mathbf{K}_t \widehat{\boldsymbol{\Sigma}}_t \mathbf{K}_t^\top\end{aligned}$$

grâce à (H₆) et au fait que les termes rectangles sont nuls : $\langle \mathbf{w}_{t+1} - \mathbf{w}_{t+1|t}, \widehat{\mathbf{e}}_t \rangle$ par les Lemmes 6.3 et 6.4 , et $\langle \mathbf{w}_t - \mathbf{w}_{t|t-1}, \mathbf{e}_{t+1} \rangle$ par les Lemmes 6.2 et 6.3.

(d) Démonstration de (T₄) : On a $\mathbf{w}_1 = \mathbf{F}_1 \mathbf{w}_0 + \mathbf{G}_1 \mathbf{e}_1$ et $\mathbf{w}_{1|0} = \mathbf{0}$ puis que $V_0 = \{0\}$.
Donc

$$\mathbf{P}_{1|0} = \|\mathbf{w}_1\|^2 = \mathbf{F}_1 \boldsymbol{\Sigma}_0 \mathbf{F}_1^\top + \mathbf{G}_1 \boldsymbol{\Sigma}_1 \mathbf{G}_1^\top$$

□

Proposition 6.7. *Supposons remplies les hypothèses du m.l.d. et de la proposition 6.6. La décomposition de Wold-Cramér du p.a. $(\mathbf{z}_t; t \geq 1)$ est alors*

$$\mathbf{z}_t = \widehat{\mathbf{e}}_t + \sum_{j=1}^{t-1} \boldsymbol{\varphi}_{jt} \widehat{\mathbf{e}}_{t-j}, \quad (6.10)$$

où les innovations $\widehat{\mathbf{e}}_t$, leur matrice de covariance $\widehat{\boldsymbol{\Sigma}}_t$ et les coefficients $\boldsymbol{\varphi}_{jt}$ s'obtiennent au moyen des relations :

$$\begin{aligned}\widehat{\boldsymbol{\Sigma}}_t &= \mathbf{H}_t \mathbf{P}_{t|t-1} \mathbf{H}_t^\top + \mathbf{S}_t, \\ \mathbf{K}_t &= \mathbf{F}_{t+1} \mathbf{P}_{t|t-1} \mathbf{H}_t^\top \widehat{\boldsymbol{\Sigma}}_t^{-1}, \\ \widehat{\mathbf{e}}_t &= \mathbf{z}_t - \mathbf{H}_t \mathbf{w}_{t|t-1}, \\ \boldsymbol{\varphi}_{tj} &= \mathbf{H}_t \left(\prod_{i=0}^{j-2} \mathbf{F}_{t-i} \right) \mathbf{K}_{t-j}, \quad j = 1, \dots, t-1,\end{aligned} \quad (6.11)$$

$$\mathbf{w}_{t+1|t} = \mathbf{F}_{t+1} \mathbf{w}_{t|t-1} + \mathbf{K}_t \widehat{\mathbf{e}}_t,$$

$$\mathbf{P}_{t+1|t} = \mathbf{F}_{t+1} \mathbf{P}_{t|t-1} \mathbf{F}_{t+1}^\top + \mathbf{G}_{t+1} \boldsymbol{\Sigma}_{t+1} \mathbf{G}_{t+1}^\top - \mathbf{K}_t \widehat{\boldsymbol{\Sigma}}_t \mathbf{K}_t^\top,$$

avec les conditions initiales

$$\begin{aligned}\mathbf{P}_{1|0} &= \mathbf{F}_1 \boldsymbol{\Sigma}_0 \mathbf{F}_1^\top + \mathbf{G}_1 \boldsymbol{\Sigma}_1 \mathbf{G}_1^\top \\ \mathbf{w}_{1|0} &= \mathbf{0}.\end{aligned}$$

Démonstration. La plupart de ces relations proviennent de la Proposition 6.6. Il reste à montrer (6.11). Partons de

$$\mathbf{z}_t = \widehat{\mathbf{e}}_t + \mathbf{H}_t \mathbf{w}_{t|t-1},$$

où

$$\mathbf{w}_{t|t-1} = \mathbf{F}_t \mathbf{w}_{t-1|t-2} + \mathbf{K}_{t-1} \widehat{\mathbf{e}}_{t-1}.$$

Par l'application de la Proposition 3.6 au processus $\{\mathbf{w}_{t|t-1}; t \geq 1\}$, avec \mathbf{F}_t au lieu de $\boldsymbol{\phi}_t$ et $\mathbf{K}_{t-1} \widehat{\mathbf{e}}_{t-1}$ au lieu de \mathbf{e}_t , il vient comme coefficient de $\widehat{\mathbf{e}}_{t-j}$, $j \geq 1$:

$$\left(\prod_{i=0}^{j-2} \mathbf{F}_{t-i} \right) \mathbf{K}_{t-j}$$

d'où (6.11). □

On peut déduire de ce qui précède un algorithme pour le calcul de la fonction de vraisemblance d'un processus ARMA(p, q) évolutif, vectoriel et gaussien. Considérons une réalisation de longueur n notée $\mathbf{z} = (z_1^\top, \dots, z_n^\top)$. Soit $\boldsymbol{\beta}$ le vecteur des paramètres. On désigne par $\boldsymbol{\Gamma}$ une matrice $n \times n$ de blocs $k \times k$, le bloc en position (t, s) étant $\boldsymbol{\Gamma}_{ts} = \langle \mathbf{z}_t, \mathbf{z}_s \rangle$, pour $t, s = 1, \dots, n$. On utilise l'algorithme de la section 3.6 pour déterminer les $\boldsymbol{\Gamma}_{ts}$ requis. Dans ces conditions, la f.v. s'écrit

$$L(\boldsymbol{\beta}; \mathbf{z}) = (2\pi)^{-kn/2} (\det \boldsymbol{\Gamma})^{-1/2} \exp\left\{-\frac{1}{2} \mathbf{z}^\top \boldsymbol{\Gamma}^{-1} \mathbf{z}\right\}.$$

On met le processus ARMA(p, q) évolutif sous la forme d'un m.l.d. à l'aide de la Proposition 3.32. Avec les notations ci-dessus, on a donc $\mathbf{a}_t = \mathbf{0}$ et donc $\mathbf{S}_t = \mathbf{0}$. Les matrices $\boldsymbol{\Sigma}_t$, \mathbf{F}_t et \mathbf{G}_t dépendent des paramètres $\boldsymbol{\beta}$ tandis que $\mathbf{H}_t = \mathbf{H}$ est une matrice constante. La matrice $\boldsymbol{\Sigma}_0$ peut être déterminée comme suit. Des expressions données pour les blocs constituant \mathbf{w}_t , on déduit que le h -ième bloc s'écrit

$$(\mathbf{w}_t)_h = \sum_{j=h}^p \boldsymbol{\varphi}_{j,t+h-1} \mathbf{z}_{t+h-1-j} - \sum_{j=h-1}^q \boldsymbol{\theta}_{j,t+h-1} \mathbf{e}_{t+h-1-j}, \quad (6.12)$$

où l'on note $\boldsymbol{\theta}_{0t} = -\mathbf{I}_k$. Dès lors, le bloc (h, ℓ) de $\boldsymbol{\Sigma}_0$ vaut

$$\begin{aligned}
\langle (\mathbf{w}_0)_h, (\mathbf{w}_0)_\ell \rangle &= \sum_{i=h}^p \sum_{j=\ell}^p \phi_{i,h-1} \Gamma_{h-1-i,\ell-1-j} \phi_{j,\ell-1}^\top \\
&\quad - \sum_{i=h}^p \sum_{j=\ell-1}^q \phi_{i,h-1} \Lambda_{h-1-i,\ell-1-j} \theta_{j,\ell-1}^\top \\
&\quad - \sum_{i=h-1}^q \sum_{j=\ell}^p \theta_{i,h-1} \Lambda_{\ell-1-j,h-1-i}^\top \phi_{j,\ell-1}^\top \\
&\quad + \sum_{i=h-1}^q \sum_{j=\ell-1}^q \theta_{i,h-1} \Sigma_{h-1-i,\ell-1-j} \theta_{j,\ell-1}^\top
\end{aligned}$$

où

$$\begin{aligned}
\Sigma_{ts} &= \begin{cases} \Sigma_t, & t = s \\ \mathbf{0}, & t \neq s \end{cases} \\
\Lambda_{ts} &= \langle \mathbf{z}_t, \tilde{\mathbf{e}}_s \rangle.
\end{aligned}$$

Il reste à déterminer Λ_{ts} pour t et s vérifiant les inégalités $-p \leq t \leq -1$ et $-q \leq s \leq 0$. Or Λ_{ts} vérifie les relations suivantes

$$\Lambda_{ts} = \begin{cases} 0, & s > t \\ \Sigma_t, & s = t \\ \sum_{j=1}^{t-s} \phi_{jt} \Lambda_{t-j,s} - \theta_{t-s} \Sigma_s, & s < t. \end{cases}$$

Si l'on suppose que ϕ_{jt}, θ_{jt} et Σ_t sont constants pour $t \leq 1$, les expressions ci-dessus se simplifient. De plus, on peut alors déterminer Λ_{ts} par récurrence en commençant par

$$\begin{aligned}
\Lambda_{00} &= \Lambda_{-1,-1} = \Lambda_{-2,-2} = \dots \\
\Lambda_{0,-1} &= \phi_{10} \Lambda_{-1,-1} - \theta_{10} \Sigma_1 = \Sigma_{-1,-2} = \dots \\
\Lambda_{0,-2} &= \phi_{10} \Lambda_{-1,-2} + \phi_{20} \Lambda_{-2,-2} - \theta_{20} \Sigma_1 = \dots
\end{aligned}$$

Cet algorithme généralise celui de Kohn et Ansley (1982) au cas évolutif. Maintenant que tous les éléments de base ont été déterminés, les innovations $\hat{\mathbf{e}}_t$ de $(\mathbf{z}_t; t \geq 1)$ sont données par application de la Proposition 6.7 (de même que les coefficients de la décomposition de Wold-Cramér, bien qu'ils ne soient pas nécessaires). Considérons le vecteur $\hat{\mathbf{e}} = (\hat{\mathbf{e}}_1^\top, \dots, \hat{\mathbf{e}}_n^\top)^\top$ et effectuons le changement de variable qui fait passer de \mathbf{z} à $\hat{\mathbf{e}}$. La matrice jacobienne déduite de (6.10) est triangulaire inférieure et contient des 1 sur la diagonale principale. Le jacobien de la transformation vaut donc 1. Or les $\hat{\mathbf{e}}_t$ sont des v.a. vectorielles mutuellement non corrélées, de vecteur moyenne $\mathbf{0}$ et de matrice de covariance $\hat{\Sigma}_t = \mathbf{H} \mathbf{P}_t \mathbf{H}^\top$, grâce au Lemme 6.5, c'est-à-dire le bloc $(1, 1)$ de la matrice $\mathbf{P}_{t|t-1}$. Puisque le processus est gaussien, les $\hat{\mathbf{e}}_t$ ont une distribution liée normale et sont donc mutuellement indépendants. La f.v. peut donc être réécrite

$$\begin{aligned}
L(\boldsymbol{\beta}; \mathbf{z}) &= \prod_{t=1}^n (2\pi)^{-k/2} [\det(\hat{\boldsymbol{\Sigma}}_t)]^{-1/2} \exp \left\{ -\frac{1}{2} \hat{\mathbf{e}}_t^\top \hat{\boldsymbol{\Sigma}}_t^{-1} \hat{\mathbf{e}}_t \right\} \\
&= (2\pi)^{-nk/2} \left[\prod_{t=1}^n \det(\hat{\boldsymbol{\Sigma}}_t) \right]^{-1/2} \exp \left\{ -\frac{1}{2} \sum_{t=1}^n \hat{\mathbf{e}}_t^\top \hat{\boldsymbol{\Sigma}}_t^{-1} \hat{\mathbf{e}}_t \right\}.
\end{aligned}$$

6.5. La méthode rapide pour les modèles ARMA constants scalaires

Proposition 6.8. *Supposons que pour tout t , \mathbf{H}_t , \mathbf{F}_t , \mathbf{G}_t , $\boldsymbol{\Sigma}_t$ et \mathbf{S}_t soient des matrices constantes notées respectivement \mathbf{H} , \mathbf{F} , \mathbf{G} , $\boldsymbol{\Sigma}$ et \mathbf{S} . Une condition suffisante pour qu'il existe un p.s.s. $(\mathbf{z}_t; t \in \mathbb{Z})$ qui soit solution du m.l.d. est*

- (a) $\det(\mathbf{I}_r - \lambda \mathbf{F}) = 0$ pour $|\lambda| \leq 1$,
- (b) $\mathbf{P}_{1|0} = \boldsymbol{\Sigma}_0$, c' est-à-dire que $\boldsymbol{\Sigma}_0$ est solution de l'équation matricielle

$$\boldsymbol{\Sigma}_0 = \mathbf{F} \boldsymbol{\Sigma}_0 \mathbf{F}^\top + \mathbf{G} \boldsymbol{\Sigma} \mathbf{G}^\top. \quad (6.13)$$

Démonstration. La décomposition de Wold-Cramér de $(\mathbf{w}_t; t \geq 0)$ donnée par (6.8) s'écrit ici

$$\mathbf{w}_t = \sum_{j=0}^{t-1} \mathbf{F}^j \mathbf{G} \mathbf{e}_{t-j} + \mathbf{F}^t \mathbf{w}_0.$$

On en déduit, pour $t \leq s$:

$$\begin{aligned}
\langle \mathbf{w}_t, \mathbf{w}_s \rangle &= \left\langle \sum_{j=0}^{t-1} \mathbf{F}^j \mathbf{G} \mathbf{e}_{t-j} + \mathbf{F}^t \mathbf{w}_0, \sum_{i=0}^{s-1} \mathbf{F}^i \mathbf{G} \mathbf{e}_{s-i} + \mathbf{F}^s \mathbf{w}_0 \right\rangle \\
&= \sum_{j=0}^{t-1} \mathbf{F}^j \mathbf{G} \boldsymbol{\Sigma} \mathbf{G} (\mathbf{F}^\top)^{s-t+j} + \mathbf{F}^t \boldsymbol{\Sigma}_0 (\mathbf{F}^\top)^s.
\end{aligned} \quad (6.14)$$

Multiplions (6.13) à gauche par \mathbf{F}^j et à droite par $(\mathbf{F}^\top)^{s-t+i}$ et sommons membre à membre pour $j = 0, \dots, t-1$. Il vient

$$\boldsymbol{\Sigma}_0 \mathbf{F}^{s-t} = \mathbf{F}^t \boldsymbol{\Sigma}_0 (\mathbf{F}^\top)^s + \sum_{j=0}^{t-1} \mathbf{F}^j \mathbf{G} \boldsymbol{\Sigma} \mathbf{G}^\top (\mathbf{F}^\top)^{s-t+j}. \quad (6.15)$$

Par comparaison de (6.14) avec (6.15), on voit que $\langle \mathbf{w}_t, \mathbf{w}_s \rangle = \boldsymbol{\Sigma}_0 \mathbf{F}^{s-t}$ ce qui ressemble à la fonction d'autocovariance d'un p.s.s. AR(1). Si la condition (a) est remplie, la Proposition 4.6 nous assure de l'existence de ce p.s.s. Considérons un b.b.s.c. $(\mathbf{a}_t; t \in \mathbb{Z})$ tel que $\mathbf{a}_t \perp \mathbf{e}_s$ pour tout t, s avec $t \neq s$. Définissons $(\mathbf{z}_t; t \in \mathbb{Z})$ à partir de (\mathbf{H}_2) pour tout $t \in \mathbb{Z}$. Puisque

$$\begin{aligned}
\langle \mathbf{z}_t, \mathbf{z}_s \rangle &= \mathbf{H} \langle \mathbf{w}_t, \mathbf{w}_s \rangle \mathbf{H}^\top + \langle \mathbf{a}_t, \mathbf{a}_s \rangle \\
&= \mathbf{H} \boldsymbol{\Sigma}_0 \mathbf{F}^{s-t} \mathbf{H}^\top + \mathbf{S} \delta_{t-s}
\end{aligned}$$

(où δ_t est le symbole de Kronecker), le p.a. $(\mathbf{z}_t; t \in \mathbb{Z})$ est un p.s.s. \square

Proposition 6.9. Soit $(z_t; t \geq 1)$ un p.a. univarié qui vérifie les hypothèses de Proposition 6.8, avec $\mathbf{a}_t = \mathbf{0}$ et $\mathbf{S} = \mathbf{0}$. Notons σ^2 au lieu de Σ_t , $\hat{\sigma}_t^2 = \mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^\top$ au lieu de $\hat{\Sigma}_t$. Dès lors (6.9) et (T_3) s'écrivent respectivement

$$\begin{aligned}\mathbf{K}_t &= \mathbf{F}\mathbf{P}_{t|t-1}\mathbf{H}^\top / \hat{\sigma}_t^2, \\ \mathbf{P}_{t+1|t} &= \mathbf{F}\mathbf{P}_{t|t-1}\mathbf{F}^\top + \sigma_t^2 \mathbf{G}\mathbf{G}^\top - \hat{\sigma}_t^2 \mathbf{K}_t \mathbf{K}_t^\top.\end{aligned}$$

Les relations énoncées dans la Proposition 6.7 peuvent être remplacées par les suivantes :

$$\alpha_{t-1} = \mathbf{H}\mathbf{L}_{t-1} / \hat{\sigma}_{t-1}^2,$$

$$\mathbf{P}_{t|t-1} = \mathbf{P}_{t-1|t-2} - \mathbf{L}_{t-1}\mathbf{L}_{t-1}^\top / \hat{\sigma}_{t-1}^2, \quad (6.16)$$

$$\mathbf{Q}_t = \mathbf{Q}_{t-1} - \alpha_{t-1} \mathbf{F}\mathbf{L}_{t-1}, \quad (6.17)$$

$$\mathbf{L}_t = \mathbf{F}\mathbf{L}_{t-1} - \alpha_{t-1} \mathbf{Q}_{t-1}, \quad (6.18)$$

$$\hat{\sigma}_t^2 = \hat{\sigma}_{t-1}^2 (1 - \alpha_{t-1}^2), \quad (6.19)$$

$$\mathbf{K}_t = \mathbf{Q}_t / \hat{\sigma}_t^2, \quad (6.20)$$

$$\hat{e}_t = z_t - \mathbf{H}\mathbf{w}_{t|t-1}, \quad (6.21)$$

$$\boldsymbol{\varphi}_{jt} = \mathbf{H}\mathbf{F}^{j-1} \mathbf{K}_{t-j}, \quad j \geq 1,$$

$$\mathbf{w}_{t+1|t} = \mathbf{F}\mathbf{w}_{t|t-1} + \mathbf{K}_t \hat{e}_t, \quad (6.22)$$

avec les conditions initiales

$$\mathbf{Q}_1 = \mathbf{L}_1 = \mathbf{F}\Sigma_0\mathbf{H}^\top,$$

$$\mathbf{P}_{1|0} = \Sigma_0,$$

$$\hat{\sigma}_1^2 = \mathbf{H}\mathbf{P}_{1|0}\mathbf{H}^\top,$$

$$\hat{e}_1 = z_1,$$

$$\mathbf{w}_{1|0} = \mathbf{0}.$$

Démonstration. Les relations de la Proposition 6.7 s'écrivent ici:

$$\mathbf{P}_{t|t-1} = \mathbf{F}\mathbf{P}_{t-1|t-2}\mathbf{F}^\top + \mathbf{G}\mathbf{G}^\top \sigma_t^2 - \mathbf{K}_{t-1}\mathbf{K}_{t-1}^\top \hat{\sigma}_{t-1}^2, \quad (6.23)$$

$$\hat{\sigma}_t^2 = \mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^\top, \quad (6.24)$$

$$\mathbf{K}_t = \mathbf{F}\mathbf{P}_{t|t-1}\mathbf{H}^\top / \hat{\sigma}_t^2, \quad (6.25)$$

$$\hat{e}_t = z_t - \mathbf{H}\mathbf{w}_{t|t-1}, \boldsymbol{\varphi}_{jt} = \mathbf{H}\mathbf{F}^{j-1} \mathbf{K}_{t-j},$$

$$\mathbf{w}_{t+1|t} = \mathbf{F}\mathbf{w}_{t|t-1} + \mathbf{K}_t \hat{e}_t,$$

avec les conditions initiales

$$\begin{aligned}\mathbf{P}_{1|0} &= \mathbf{F}\boldsymbol{\Sigma}_0\mathbf{F}^\top + \mathbf{G}\mathbf{G}^\top\sigma^2, \\ \mathbf{w}_{1|0} &= \mathbf{0}.\end{aligned}$$

Montrons que si les relations exprimées dans la proposition sont valables au temps t , elle le sont encore au temps $t+1$. Notons $\alpha_s = \mathbf{H}\mathbf{L}_s/\hat{\sigma}_s^2 = \mathbf{L}_s^\top\mathbf{H}^\top/\hat{\sigma}_s^2$ et $\mathbf{Q}_s = \hat{\sigma}_s^2\mathbf{K}_s$ pour tout s . Il suffit donc de montrer que

$$\mathbf{P}_{t+1|t} = \mathbf{P}_{t|t-1} - \mathbf{L}_t\mathbf{L}_t^\top/\hat{\sigma}_t^2, \quad (6.26)$$

$$\mathbf{Q}_{t+1} = \mathbf{Q}_t - \alpha_t\mathbf{F}\mathbf{L}_t, \quad (6.27)$$

$$\hat{\sigma}_{t+1}^2 = \hat{\sigma}_t^2(1 - \alpha_t^2), \quad (6.28)$$

pour que les autres relations soient automatiquement valables. Notons qu'on n'a pas besoin d'établir la relation relative à \mathbf{L}_{t+1} qui est une définition alors que la relation donnant $\mathbf{P}_{t+1|t}$ ne sert que dans la démonstration mais peut être omise de l'algorithme. Partons de l'expression de $\mathbf{P}_{t+1|t}$ tirée de la Proposition 6.7 :

$$\begin{aligned}\mathbf{P}_{t+1|t} &= \mathbf{F}\mathbf{P}_{t|t-1}\mathbf{F}^\top + \mathbf{G}\mathbf{G}^\top\sigma^2 - \mathbf{K}_t\mathbf{K}_t^\top\hat{\sigma}_t^2 \\ &= \mathbf{F}(\mathbf{P}_{t-1|t-2} - \mathbf{L}_{t-1}\mathbf{L}_{t-1}^\top/\hat{\sigma}_{t-1}^2)\mathbf{F}^\top + \mathbf{G}\mathbf{G}^\top\sigma^2 - \mathbf{Q}_t\mathbf{Q}_t^\top/\hat{\sigma}_t^2\end{aligned}$$

grâce à (6.16) et à (6.20) . En ajoutant et retranchant $\mathbf{Q}_{t-1}\mathbf{Q}_{t-1}^\top/\hat{\sigma}_{t-1}^2$ on obtient après groupement et prise en compte de (6.23) et (6.19)

$$\begin{aligned}\mathbf{P}_{t+1|t} &= \mathbf{F}\mathbf{P}_{t-1|t-2}\mathbf{F}^\top + \mathbf{G}\mathbf{G}^\top\sigma^2 - \mathbf{Q}_{t-1}\mathbf{Q}_{t-1}^\top/\hat{\sigma}_{t-1}^2 \\ &\quad + \mathbf{Q}_{t-1}\mathbf{Q}_{t-1}^\top/\hat{\sigma}_{t-1}^2 - \mathbf{F}\mathbf{L}_{t-1}\mathbf{L}_{t-1}^\top\mathbf{F}^\top/\hat{\sigma}_{t-1}^2 - \mathbf{Q}_t\mathbf{Q}_t^\top/\hat{\sigma}_t^2 \\ &= \mathbf{P}_{t|t-1} + (1/\hat{\sigma}_t^2)\{(1 - \alpha_{t-1}^2)(\mathbf{Q}_{t-1}\mathbf{Q}_{t-1}^\top - \mathbf{F}\mathbf{L}_{t-1}\mathbf{L}_{t-1}^\top\mathbf{F}^\top) - \mathbf{Q}_t\mathbf{Q}_t^\top\} \\ &= \mathbf{P}_{t|t-1} + (1/\hat{\sigma}_t^2)\{(1 - \alpha_{t-1}^2)(\mathbf{Q}_{t-1}\mathbf{Q}_{t-1}^\top - \mathbf{F}\mathbf{L}_{t-1}\mathbf{L}_{t-1}^\top\mathbf{F}^\top) \\ &\quad - (\mathbf{Q}_{t-1} - \alpha_{t-1}\mathbf{F}\mathbf{L}_{t-1})(\mathbf{Q}_{t-1} - \alpha_{t-1}\mathbf{F}\mathbf{L}_{t-1})^\top\}\end{aligned}$$

grâce à (6.17) . En effectuant, il vient

$$\begin{aligned}\mathbf{P}_{t+1|t} &= \mathbf{P}_{t|t-1} - (1/\hat{\sigma}_t^2)\{\alpha_{t-1}^2\mathbf{Q}_{t-1}\mathbf{Q}_{t-1}^\top + \mathbf{F}\mathbf{L}_{t-1}\mathbf{L}_{t-1}^\top\mathbf{F}^\top - \alpha_{t-1}\mathbf{Q}_{t-1}\mathbf{L}_{t-1}^\top\mathbf{F}^\top \\ &\quad - \alpha_{t-1}\mathbf{F}\mathbf{L}_{t-1}\mathbf{Q}_{t-1}^\top\} \\ &= \mathbf{P}_{t|t-1} - (\mathbf{F}\mathbf{L}_{t-1} - \alpha_{t-1}\mathbf{Q}_{t-1})(\mathbf{F}\mathbf{L}_{t-1} - \alpha_{t-1}\mathbf{Q}_{t-1})^\top/\hat{\sigma}_t^2 \\ &= \mathbf{P}_{t|t-1} - \mathbf{L}_t\mathbf{L}_t^\top/\hat{\sigma}_t^2\end{aligned}$$

après avoir exploité (6.18) . Ceci établit (6.26) . On démontre (6.27) comme suit :

$$\begin{aligned}
\mathbf{Q}_{t+1} &= \hat{\sigma}_{t+1}^2 \mathbf{K}_{t+1} \\
&= \mathbf{F} \mathbf{P}_{t+1|t} \mathbf{H}^\top \\
&= \mathbf{F} (\mathbf{P}_{t|t-1} - \mathbf{L}_t \mathbf{L}_t^\top \hat{\sigma}_t^2) \mathbf{H}^\top \\
&= \mathbf{F} \mathbf{P}_{t|t-1} \mathbf{H}^\top - \alpha_{t-1} \mathbf{F} \mathbf{L}_t \\
&= \hat{\sigma}_t^2 \mathbf{K}_t - \alpha_{t-1} \mathbf{F} \mathbf{L}_t \\
&= \mathbf{Q}_t - \alpha_t \mathbf{F} \mathbf{L}_t
\end{aligned}$$

en utilisant successivement (6.25) , (6.26) , (6.25) à nouveau et (6.20) . Ensuite, il vient de (6.21)

$$\begin{aligned}
\hat{\sigma}_{t+1}^2 &= \mathbf{H} \mathbf{P}_{t+1|t} \mathbf{H}^\top \\
&= \mathbf{H} (\mathbf{P}_{t|t-1} - \mathbf{L}_t \mathbf{L}_t^\top) \mathbf{H}^\top \\
&= \hat{\sigma}_t^2 - (\mathbf{H} \mathbf{L}_t)^2 / \hat{\sigma}_t^2 \\
&= \hat{\sigma}_t^2 (1 - \alpha_t^2),
\end{aligned}$$

ce qui démontre (6.28) . Il reste donc à établir la validité des conditions initiales. Les valeurs de $\hat{\sigma}_1^2$, $\hat{\varepsilon}_1$ et $\mathbf{w}_{1|0}$ viennent directement de la Proposition 6.7. On veut $\mathbf{K}_1 = \mathbf{F} \mathbf{P}_{1|0} \mathbf{H}^\top / \hat{\sigma}_1^2$ en vertu de (6.25) , donc $\mathbf{Q}_1 = \hat{\sigma}_1^2 \mathbf{K}_1 = \mathbf{F} \mathbf{P}_{1|0} \mathbf{H}^\top = \mathbf{F} \boldsymbol{\Sigma}_0 \mathbf{H}^\top$, compte tenu de la Proposition 6.8 . Finalement, par le même argument

$$\begin{aligned}
\mathbf{P}_{2|1} - \mathbf{P}_{1|0} &= \mathbf{F} \mathbf{P}_{1|0} \mathbf{F}^\top + \mathbf{G} \mathbf{G}^\top \sigma^2 - \mathbf{K}_1 \mathbf{K}_1^\top \hat{\sigma}_1^2 - \mathbf{P}_{1|0} \\
&= -\mathbf{K}_1 \mathbf{K}_1^\top \hat{\sigma}_1^2 \\
&= -\mathbf{Q}_1 \mathbf{Q}_1^\top / \hat{\sigma}_1^2.
\end{aligned}$$

On choisit donc $\mathbf{L}_1 = \mathbf{Q}_1 = \mathbf{F} \boldsymbol{\Sigma}_0 \mathbf{H}^\top$ pour satisfaire (6.26) en $t = 1$. Mais alors

$$\begin{aligned}
\mathbf{Q}_2 &= \hat{\sigma}_2^2 \mathbf{K}_2 = \mathbf{F} \mathbf{P}_{2|1} \mathbf{H}^\top = \mathbf{F} (\mathbf{P}_{1|0} - \mathbf{L}_1 \mathbf{L}_1^\top / \hat{\sigma}_1^2) \mathbf{H}^\top \\
&= \mathbf{F} \mathbf{P}_{1|0} \mathbf{H}^\top - \mathbf{F} \mathbf{L}_1 \mathbf{L}_1^\top \mathbf{H}^\top / \hat{\sigma}_1^2 \\
&= \mathbf{F} \boldsymbol{\Sigma}_0 \mathbf{H}^\top - \alpha_1 \mathbf{F} \mathbf{L}_1 \\
&= \mathbf{Q}_1 - \alpha_1 \mathbf{F} \mathbf{L}_1,
\end{aligned}$$

ce qui vérifie (6.27) en $t = 1$. □

Proposition 6.10. Soit $(z_t; t \in \mathbb{Z})$ un processus ARMA(p, q) constant sous forme d'un m.l.d. comme dans la Proposition 6.9 , avec $r = \max(p, q + 1)$. Si $p > q$, les équations correspondant aux lignes $j = q + 1, \dots, r$ de (6.22) peuvent être remplacées, pour $t \geq p - q + 1$, par

$$(\mathbf{w}_{t+1|t})_j = (\mathbf{w}_{t|t-1})_{j+1} + \phi_j z_t.$$

Démonstration. Puisque le processus ARMA(p, q) est univarié et constant, les matrices \mathbf{H} , \mathbf{F} et \mathbf{G} de la Proposition 3.32 s'écrivent ici

$$\mathbf{H} = (1, 0, \dots, 0), \quad \mathbf{F} = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_r \\ 0 \dots 0 \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} 1 \\ -\theta_1 \\ -\theta_2 \\ \vdots \\ -\theta_{r-1} \end{bmatrix}.$$

Dès lors, (6.21) implique $z_t = \hat{e}_t + (\mathbf{w}_{t|t-1})_1$ tandis que (6.22) peut s'écrire

$$\begin{aligned} (\mathbf{w}_{t|t-1})_1 &= \phi_1 (\mathbf{w}_{t-1|t-2})_1 + (\mathbf{K}_{t-1})_1 \hat{e}_{t-1} + (\mathbf{w}_{t-1|t-2})_2 \\ (\mathbf{w}_{t-1|t-2})_2 &= \phi_2 (\mathbf{w}_{t-2|t-3})_1 + (\mathbf{K}_{t-2})_2 \hat{e}_{t-2} + (\mathbf{w}_{t-2|t-3})_3 \\ &\dots \\ (\mathbf{w}_{t-r+1|t-r})_r &= \phi_r (\mathbf{w}_{t-r|t-r-1})_1 + (\mathbf{K}_{t-r})_r \hat{e}_{t-r}, \end{aligned}$$

où

$$(\mathbf{w}_{t-j|t-j-1})_1 = z_{t-j} - \hat{e}_{t-j}.$$

On en déduit donc que

$$z_t = \hat{e}_t + \sum_{j=1}^{p(t)} \phi_j (z_{t-j} - \hat{e}_{t-j}) + \sum_{j=1}^{q(t)} (\mathbf{K}_{t-j})_j \hat{e}_{t-j},$$

où on a posé $p(t) = \min(t-2, p)$ et $q(t) = \min(t-1, r)$, tenant compte de ce que $\phi_j = 0$ pour $j > p$ et $\mathbf{w}_{1|0} = \mathbf{0}$. Définissons

$$x_t = z_t - \sum_{j=1}^{p(t)} \phi_j z_{t-j} \quad \text{pour} \quad t \geq 1.$$

Puisque les \hat{e}_t sont les innovations de $(z_t; t \geq 1)$, ce sont aussi les innovations de $(x_t; t \geq 1)$ (Proposition 2.38, remarque 2). Il s'ensuit que

$$x_t = \hat{e}_t - \sum_{j=1}^{p(t)} \phi_j \hat{e}_{t-j} + \sum_{j=1}^{q(t)} (\mathbf{K}_{t-j})_j \hat{e}_{t-j}$$

est la décomposition de Wold-Cramér du processus $(x_t; t \geq 1)$. Pour $t \geq p+2$, on a

$$x_t = z_t - \sum_{j=1}^p \phi_j z_{t-j} = e_t - \sum_{j=1}^q \theta_j e_{t-j}$$

donc le processus est q -dépendant. Sa décomposition de Wold-Cramér, qui est unique, est donc telle que les termes pour $j > q$ sont nuls, ce qui entraîne que $(\mathbf{K}_{t-j})_j = \phi_j$ pour $j > q$ et $t \geq p+2$. Pour $j = q+1$ cette simplification apparaît dès que $t \geq p-q+1$, de sorte que $(\mathbf{K}_t)_j = \phi_j$ pour $j = q+1, \dots, r$. La relation (6.22) devient alors

$$\begin{aligned} (\mathbf{w}_{t+1|t})_j &= \phi_j(\mathbf{w}_{t|t-1})_1 + (\mathbf{w}_{t|t-1})_{j+1} + \phi_j \hat{e}_t \\ &= (\mathbf{w}_{t|t-1})_{j+1} + \phi_j z_t. \end{aligned}$$

□

Proposition 6.11. Les composantes μ_i de $\mu = \mathbf{\Sigma}_0 \mathbf{H}^\top$ peuvent s'obtenir par les équations

$$\mu_h = \sum_{j=h}^r (\phi_j \gamma_{j-h+1} - \theta_{j-1} \lambda_{j-h}), \quad h = 1, \dots, r, \quad (6.28bis)$$

où

$$\lambda_i = -\theta_i \sigma^2 + \sum_{j=1}^{\min(p,i)} \phi_j \lambda_{i-j}, \quad i = 0, \dots, r, \quad (6.29)$$

et donc

$$(\mathbf{Q}_1)_i = (\mathbf{L}_1)_i = \phi_i \mu_1 + \mu_{i+1}, \quad i = 1, \dots, r, \quad (6.30)$$

avec $\mu_{r+1} = 0$.

Démonstration. On peut récrire (6.12) sous la forme

$$(\mathbf{w}_t)_h = \sum_{j=h}^r (\phi_j z_{t-j+h-1} - \theta_{j-1} e_{t-j+h}).$$

Puisque $(\mathbf{w}_t)_1 = z_t$, il vient en notant $\lambda_i = \langle z_t, e_{t-i} \rangle$:

$$\begin{aligned} \mu_h &= (\mathbf{\Sigma}_0)_{h1} = \langle (\mathbf{w}_t)_h, (\mathbf{w}_t)_1 \rangle \\ &= \left\langle \sum_{j=h}^r (\phi_j z_{t-j+h-1} - \theta_{j-1} e_{t-j+h}), z_t \right\rangle \\ &= \sum_{j=h}^r (\phi_j \gamma_{j-h+1} - \theta_{j-1} \lambda_{j-h}). \end{aligned}$$

Or (6.4) entraîne (6.29). Considérons maintenant $\mathbf{Q}_1 = \mathbf{L}_1 = \mathbf{F} \mathbf{\Sigma}_0 \mathbf{H}^\top = \mathbf{F} \mu$. De par la forme prise par \mathbf{F} , il vient (6.30). □

Des propositions de ce paragraphe on peut déduire un algorithme rapide de la f.v. d'un processus ARMA univarié constant. Il s'agit à nouveau d'évaluer (6.7) où les innovations $h_t e_t$ sont à remplacer par \hat{e}_t et leur écart type $h_t \sigma$ est à remplacer par $\hat{\sigma}_t$. Par conséquent e_t est à remplacer par $(\sigma \hat{e}_t / \hat{\sigma}_t)$ dans les notations de ce paragraphe-ci. La méthode consiste en les étapes suivantes :

- 1) calcul des rapports γ_j / σ^2 pour $j = 0, \dots, r$ au moyen de l'algorithme de la section 4.13, pour les ϕ_i ($i = 1, \dots, p$) et θ_i , ($i = 1, \dots, q$) donnés;
- 2) calcul des rapports λ_i / σ^2 et μ_h / σ^2 par les relations (6.28bis) et (6.29) ;

- 3) calcul des rapports Q_1/σ^2 et L_1/σ^2 au moyen de (6.30) et de $\hat{\sigma}_1^2/\sigma^2 = \mu_1/\sigma^2$;
- 4) on pose $\hat{e}_1 = z_1$ et $\mathbf{w}_{1|0} = 0$ puis on calcule les \hat{e}_t par récurrence en utilisant les relations (6.17), (6.18), (6.19), (6.20), (6.21) et (6.22), où les deux membres des trois premières sont divisés par σ^2 ;
- 5) simultanément, on a calculé $h_t^2 = \hat{\sigma}_t^2/\sigma^2$, $t = 1, \dots, n$, dont on déduit $\prod_{t=1}^n h_t$.

Un programme en Fortran est disponible à cette fin (Mélard, 1984).

6.6. Comparaison des méthodes d'évaluation de la fonction de vraisemblance

Le tableau qui suit porte sur les principaux algorithmes proposés pour évaluer la fonction de vraisemblance des processus ARMA(p, q). Les deux critères utilisés sont

- (i) le nombre de multiplications et de divisions;
- (ii) le nombre de mots de mémoire de zones de travail.

Seuls les termes principaux par rapport à n , p et q sont conservés. De plus, p et q sont négligés devant n et 1 est négligé devant p ou q .

Des expériences pratiques rapportées par Mélard (1984) confirment l'efficacité de l'algorithme de la section 6.5. En pratique, cet algorithme peut être modifié afin d'exploiter la convergence des $(\mathbf{K}_t)_j$ vers ϕ_j , ce qui peut fournir une approximation aussi proche que l'on veut de la vraie valeur de la f.v. en un temps plus court que la méthode de la s.c.n., voisin même du temps utilisé par la méthode de la s.c.c. Or, la méthode de la s.c.n. est une méthode approchée, au comportement incertain, surtout lorsque des zéros des polynômes AR ou MA sont proches de la frontière de stationnarité ou d'inversibilité, respectivement. Il n'y a donc plus aucune raison de l'employer.

La discussion qui précède vaut seulement pour les processus ARMA constants. Pour ce qui concerne les processus ARMA évolutifs, un seul algorithme est disponible. Cependant, l'algorithme basé sur le filtre de Kalman proposé à la section 6.4, particulièrement au processus univarié, devrait être équivalent.

Tableau 6.1 Comparaison des principaux algorithmes d'évaluation de la f.v. Notations: $r = \max(p, q + 1)$, $s = \min(p, q)$, $m = \max(p, q)$, $v =$ horizon de la prévision rétrospective.

Algorithmes	Références	Nombre de multiplications et divisions	Nombre de mots de mémoire	Remarques
Approchés				
M. de la s.c.c. (MCC)	Box&Jenkins (1976)	$n(p+q)$	0	avec une seule iteration
M. de la s.c.n. (MCN)	Box&Jenkins (1976)	$(2n+2v)(p+q)$	$2v$	
Exacts (ARMA constants)				
M. directe (inversion de Γ)		$n^3/2$	n^2	
M. de Levinson	Khobie- Zeitoune (1980)	$n^2/2$	n^2	
Newbold (1974)		$n^2/2$	n^2	
Ljung et Box (1979)		$m^3 + p^3/2$	$2m^2$	
Factorisation de Cholesky de matrice bande	Ansley (1979) Mélard (1982a)	$\frac{1}{2}p^3 + n(p + \frac{1}{2}q^2)$	m^2	
Filtre de Kalman	Gardner, Harvey & Phillips (1980)	$\frac{1}{2}p^2r^4 + n(p + \frac{1}{2}r^2)$	$\frac{1}{8}r^4$	
M. rapide	Pearlman (1980) Mélard (1984)	$(p^2 + \frac{1}{2}q^2 + 2pm + qs + \frac{1}{2}m^2) + n(p + 3q + s)$	$3(r+1)$	avec l'algorithme de Wilson (1979)
Exacts (ARMA évolutifs)				
Factorisation de Cholesky de matrice bande	Mélard (1982a)	$p^3 + n(p + q^2)$	$m(p+q) + 2m^2$	

Chapitre 7

LE MODÈLE ARIMA GÉNÉRALISÉ

Dans les chapitres précédents, une large part a été laissée aux processus ARMA évolutifs où les coefficients varient dans le temps de façon quelconque, ce qui est trop général sur le plan pratique. En revanche, on a toujours supposé que le p.a. est centré, hypothèse trop restrictive pour les applications, et que les innovations sont distribuées selon une loi normale. Le modèle ARIMA généralisé, fruit de la synthèse d'apports extérieurs et de contributions personnelles, réunit plusieurs voies de généralisation des modèles ARIMA, basées sur des principes simples. Afin de pouvoir représenter des s.c. le modèle doit être

1. suffisamment général;
2. paramétré au moyen d'un nombre limité de paramètres susceptibles d'être interprétés;
3. exprimable sous forme d'un algorithme de calcul des résidus e_t à partir des données z_t .

Un logiciel a été mis au point afin d'estimer les paramètres de ce modèle. Sous le nom de ANSECH (ANalyse des SERIES CHronologiques) (Mélard, 1982b, 1983), il a été développé pour être utilisé sur les ordinateurs Control Data (séries 6 000, Cyber 70 et 170). Pour un exemple d'illustration des différentes composantes du modèle ARIMA généralisé, nous renvoyons le lecteur à Mélard (1985) qui utilise le même exemple que Kashyap et Rao (1976).

7.1. Le modèle ARIMA généralisé

Le modèle ARIMA généralisé que nous étudions ici est défini par les trois équations suivantes:

- (le sous-modèle de la variable)

$$\frac{\nabla^d \nabla_s^D \{ \tilde{C}_\lambda(z_t - y_t^I) / f_t \} - (y_t^D + \mu_t + m_t)}{y_t^V} = w_t, \quad (7.1)$$

- (le sous-modèle ARMA évolutif)

$$w_t - \sum_{i=1}^p \phi_{it} w_{t-i} = a_t - \sum_{j=1}^q \theta_{jt} a_{t-j}, \quad (7.2)$$

- (le sous-modèle de l'innovation)

$$a_t = g_t \{y_t^S e_t + y_t^M\}, \quad (7.3)$$

dans lesquelles les notations suivantes sont utilisées :

- $(z_t; t \in \mathbb{Z})$ est le processus étudié, qui n'est pas nécessairement un p.a.s ;
- $(w_t; t \in \mathbb{Z})$ est un p.a.s. qui n'est pas nécessairement un p.s.s.;
- $(a_t; t \in \mathbb{Z})$ est le processus innovation, constitué de v.a. indépendantes mais qui n'est pas nécessairement de moyenne nulle ni stationnaire;
- $(e_t; t \in \mathbb{Z})$ est le processus innovation normalisé qui est b.b.s.c. de moyenne 0 et de variance σ^2 ;
- $y_t^I, f_t, y_t^D, \mu_t, m_t, \phi_{it}$ ($i = 1, \dots, p$), θ_{jt} ($j = 1, \dots, q$), y_t^S, g_t et y_t^M sont des fonctions déterministes de t , de forme analytique déterminée mais dépendant d'un nombre fini de paramètres inconnus;
- ∇^d est l'opérateur de différence ordinaire, itéré d fois;
- ∇_s^D est l'opérateur de différence saisonnière, de période s , itéré D fois;
- $\tilde{C}_\lambda(\cdot)$ est une fonction dépendant de paramètres λ inconnus.

On suppose que les données sont transformées en innovations normalisées e_t par la succession des opérations suivantes:

1. on soustrait y_t^I , fonction nulle partout sauf là où l'on désire représenter l'effet d'interventions (voir-ci-dessous) ;
2. on effectue une transformation instantanée $\tilde{C}_\lambda(\cdot)$;
3. on divise par la tendance multiplicative f_t ;
4. on applique les différences afin d'obtenir un p.a.s.;
5. on soustrait y_t^D , qui est une autre fonction d'intervention, capable de représenter les changements brusques de niveau;
6. on soustrait une tendance déterministe μ_t , fonction polynomiale de t ;
7. on soustrait une composante saisonnière déterministe m_t , qui est une fonction périodique de t , de somme nulle sur un intervalle de longueur égale à une période;
8. on applique le filtre ARMA évolutif ;
9. on divise les résidus par g_t qui représente une tendance en écart type des innovations ;
10. on soustrait y_t^M qui est une fonction d'intervention portant sur la moyenne des innovations;
11. on divise par y_t^S , où y_t^S est également une fonction d'intervention, qui agit donc sur l'écart type des innovations.

Remarque. Les fonctions f_t, g_t, y_t^S et y_t^M sont supposés strictement positives.

Exemples.

1. y_t^I prend en compte les interventions, les changements de définition de la grandeur étudiée, les valeurs aberrantes;
2. $\tilde{C}_\lambda(\cdot)$ peut être la transformation logarithmique ou une transformation de puissance;
3. une tendance multiplicative peut apparaître par l'influence de l'évolution démographique ou des prix ;
4. l'emploi de différences permet de représenter des séries qui n'ont pas de moyenne ni de variance;
5. la fonction d'intervention y_t^D est pratique pour modéliser un changement de niveau dans la série;
6. si la tendance est polynomiale, il sera plus facile de l'exprimer directement sous la forme de μ_t plutôt que de l'éliminer par différences ordinaires;
7. de même, si le saisonnier s'exprime sous forme d'une fonction périodique m_t , l'inclure dans le modèle évitera de devoir l'éliminer par différence saisonnière ;
8. le filtre ARMA évolutif ou le filtre ARMA constant permet de calculer les innovations à partir des observations passées, préalablement transformées, et des innovations passées ;
9. la division par g_t stabilise la variance des innovations en éliminant une tendance éventuelle dans cette variance ;
10. la fonction d'intervention y_t^M permet d'agir directement au niveau de la moyenne des innovations qui est donc susceptible d'être non nulle;
11. la fonction d'intervention y_t^S a le même effet au niveau de l'écart type des innovations, afin de représenter, par exemple, une période troublée au cours de laquelle la dispersion a été plus grande.

Il est évidemment inconcevable de réunir simultanément toutes ces extensions dans l'analyse des données d'une série. Toutefois, une ou deux d'entre elles peut s'avérer utile, voire indispensable.

7.2. Les interventions

L'approche habituelle pour traiter les interventions est de les représenter sous la forme d'une fonction y_t^I avec

$$y_t^I = \sum_{i=1}^I y_t^{(i)} \quad (7.4)$$

avec $y_t^{(i)} = v^{(i)}(B)\xi_t^{(i)}$, où les $\xi_t^{(i)}$ constituent des fonctions déterministes, par exemple à valeur binaires pour indiquer la présence d'intervention ou l'absence. Les séries $v^{(i)}(B)$ dépendent de paramètres à estimer et peuvent donc être plus parcimonieusement représentées par le rapport de deux polynômes $\omega_i(B)/\delta_i(B)$.

Cette approche est extrêmement intéressante pour modéliser l'effet d'interventions réelles sur la grandeur étudiée (changement de définition, de réglementation, de

circonstances,...), voir par exemple Box et Tiao (1975), Bhattacharyya et Layton (1979), Roy et Pellerin (1982), pourvu que ces effets puissent être décrits simplement. Quand le but est de corriger les données, afin de diminuer l'influence de phénomènes passagers (crises, grèves,...) sur les paramètres du modèle stochastique ou de représenter des effets de nature complexe, la méthode basée sur les fonctions de transfert est à la fois peu pratique et peu efficiente. En effet, pour chaque phénomène à faire intervenir, il faut introduire une s.c. $\xi_r^{(i)}$ et deux polynômes $\omega_i(B)$ et $\delta_i(B)$. Puisque $\xi_r^{(i)}$ est souvent presque partout nulle, la plupart des calculs sont effectués sur des zéros. De plus, la représentation d'interventions complexes soulève des difficultés dans le choix de ces s.c. et des polynômes.

Une approche plus appropriée pour traiter ces cas consiste à représenter y_t^I comme fonction linéaire par morceaux du type (7.4). Notons $\beta_0^{(i)}, \beta_1^{(i)}, \dots, \beta_{m_i}^{(i)}$ des réels (constantes ou paramètres) et $t_0^{(i)} \leq t_1^{(i)} \leq \dots \leq t_{m_i}^{(i)}$ des instants fixés, pour $i = 1, \dots, I$. On pose

$$y_t^I = \begin{cases} \beta_0^{(i)} + \beta_1^{(i)}(t - t_0^{(i)}), & t_0^{(i)} \leq t \leq t_1^{(i)} \\ \beta_0^{(i)} + \sum_{\ell=1}^{k-1} \beta_\ell^{(i)}(t_\ell^{(i)} - t_{\ell-1}^{(i)}) + \beta_k^{(i)}(t - t_{k-1}^{(i)}), & t_{k-1}^{(i)} \leq t \leq t_k^{(i)}, (k = 2, \dots, m_i) \\ 0 & \text{ailleurs,} \end{cases} \quad (7.4')$$

corrigeant ainsi (Mélard, 1981). De plus, en permettant l'usage des fonctions y_t^D et y_t^M dans le modèle ARIMA généralisé, on dispose de trois représentations entre lesquelles on peut choisir ou que l'on peut combiner:

$$\begin{aligned} \phi(B)\nabla(z_t - y_t^I) &= \theta(B)e_t \\ \phi(B)(\nabla z_t - y_t^D) &= \theta(B)e_t \\ \phi(B)\nabla z_t &= \theta(B)(e_t + y_t^M). \end{aligned}$$

Cette dernière représentation est appelée intervention sur l'innovation (Fox, 1972, Chang et Tiao, 1983 et Bell, 1983). Formellement, ces trois approches sont équivalentes puisque y_t^I, y_t^D et y_t^M sont liées par les relations

$$y_t^D = \nabla y_t^I \quad \text{et} \quad \phi(B)y_t^D = \theta(B)y_t^M.$$

Du point de vue de l'estimation des paramètres, il n'y a pas équivalence entre l'emploi de y_t^D ou y_t^I , d'une part, et de y_t^M , d'autre part. En effet, la présence de y_t^M introduit une modification dans l'évaluation de la f.v. qui sera traitée à la section 7.9.

Les types d'interventions discutées auparavant ont un effet sur la moyenne. Il est aussi utile de disposer d'intervention sur l'écart type, sous la forme d'une fonction y_t^S dans le modèle (Mélard, 1981):

$$\phi(B)\nabla z_t = \theta(B)(y_t^S e_t).$$

Il s'agit clairement d'un modèle ARIMAG, défini à la section 3.5. Dans ce cas, la f.v. exacte doit être évaluée au moyen d'un algorithme pour modèles ARMA évolutifs, même s'il n'y a qu'un seul instant d'intervention.

7.3. Les transformations

La famille de transformations qui est utilisée le plus fréquemment est celle des puissances

$$\begin{cases} (z_t + c)^\lambda, & \lambda \neq 0 \\ \log(z_t + c), & \lambda = 0, \end{cases}$$

en supposant $(z_t + c) > 0$ pour tout t . Afin que la famille soit continue comme fonction de λ , même en $\lambda = 0$, on définit plutôt

$$\tilde{C}_\lambda(z_t) = \begin{cases} \frac{(z_t + c)^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \log(z_t + c), & \lambda = 0. \end{cases} \quad (7.5)$$

On suppose que z_t peut être approché par un processus gaussien en négligeant le fait que le domaine de définition est borné inférieurement si $\lambda > 0$ ou supérieurement si $\lambda < 0$. Box et Cox (1964) et Ansley, Spivey et Wroblewski (1977) ont exposé une méthode pour l'estimation simultanée de λ et des autres paramètres du modèle. Certains logiciels nécessitent encore d'essayer le modèle pour quelques valeurs de λ en laissant à l'utilisateur le choix de la valeur finale. En ce qui concerne c , on sait (par exemple, Hill, 1963) que l'estimateur du m.v. vaut $-\min(z_1, \dots, z_n)$. Il est donc situé sur la frontière de l'espace des paramètres puisque $z_t + c = 0$ pour au moins un t . De plus, la f.v. est non bornée. Il convient donc d'éviter l'estimation de c . A la section 7.8, on développe la méthode d'estimation simultanée, c exclus, en utilisant la transformation sous forme normalisée :

$$C_\lambda(z_t) = \begin{cases} \frac{z_t^\lambda - 1}{\lambda G^{\lambda-1}}, & \lambda \neq 0 \\ G \log(z_t), & \lambda = 0. \end{cases} \quad (7.6)$$

où $G = (\prod_{t=1}^n z_t)^{1/n}$ est la moyenne géométrique des données. Signalons déjà que les facteurs introduits dans (7.6) proviennent de l'évaluation du jacobien de la transformation, tel qu'il résulte de (7.5).

7.4. La tendance multiplicative et la tendance en écart type des innovations

La tendance multiplicative, f_t , et la tendance en écart type des innovations, g_t , sont introduites afin d'obtenir un modèle FARIMAG (Définition 3.15). On suppose $f_t > 0$ et $g_t > 0$. Ces deux fonctions dépendent d'un petit nombre de paramètres. Compte tenu de ce que la tendance en dispersion est généralement très modérée, une fonction linéaire ou une fonction exponentielle donnent souvent des résultats suffisants, presque identiques d'ailleurs.

Il n'y a pas intérêt à inclure simultanément f_t et g_t . Dans le cas particulier d'un processus FARMA avec $f_t = \exp(\beta t)$, donc tel que $f_{t-j} = f_t \exp(-\beta j)$, on peut réécrire l'équation d'un modèle FARMA

$$f_t^{-1} \sum_{j=0}^p \phi_j \exp(\beta j) z_{t-j} = \sum_{j=0}^q \theta_j e_{t-j}$$

sous la forme d'un modèle ARMAG

$$\sum_{j=0}^p \phi'_j z_{t-j} = \sum_{j=0}^q \theta'_j (g_t e_{t-j})$$

avec

$$\phi'_j = \phi_j \exp(\beta j), \quad \theta'_j = \theta_j \exp(\beta j), \quad g_t = f_t$$

(Kiehm, Lefèvre et Mélard, 1980). Ceci montre qu'il peut y avoir confusion entre les effets de f_t et de g_t . On verra ultérieurement qu'il y a avantage à normaliser f_t et g_t de sorte que

$$\prod_{t=h}^n f_t = \prod_{t=h}^n g_t = 1$$

où $h = d + Ds + 1$.

7.5. Le sous-modèle ARMA évolutif

Le sous-modèle ARMA évolutif est un cas particulier du modèle étudié au chapitre 3, dans lequel on veille à employer un nombre limité de paramètres. Le but est souvent de disposer d'un modèle qui évolue lentement avec le temps au lieu de rester constant. Dès lors, il est acceptable de se limiter, pour les coefficients ϕ_{it} et θ_{jt} à des fonctions linéaires (ou à des fonctions exponentielles) du temps. Cela permet d'interpréter aisément le caractère évolutif du modèle et de simplifier au maximum l'évaluation de ces coefficients qui dépendent du temps. En considérant ϕ_{it} à titre d'exemple, on a donc la paramétrisation suivante:

$$\phi_{it} = \phi'_i + \phi''_i (t - v)$$

ou

$$\phi_{it} = \phi_i' \exp[\phi_i''(t - \nu)]$$

avec $\nu = d + Ds + (n + 1)/2$. Nous n'avons pas envisagé dans le logiciel d'autres possibilités telles que les coefficients variant de façon périodique (Pagano, 1978) ou par saut (Bagshaw et Johnson, 1977).

Le sous-modèle ARMA évolutif contient évidemment le sous-modèle ARMA constant comme cas particulier. Celui-ci est paramétré de façon multiplicative, c'est-à-dire

$$\phi(B) = \phi_O(B)\phi_S(B), \quad \theta(B) = \theta_O(B)\theta_S(B),$$

(les indices O et S représentent respectivement ordinaire et saisonnier).

Une autre paramétrisation du sous-modèle ARMA est disponible afin de couvrir les modèles particuliers sous-jacents au lissage exponentiel d'ordre k , au lissage exponentiel avec tendance, au lissage exponentiel de Holt-Winters (Mélard et Rouland, 1986).

7.6. La différence ordinaire et la tendance déterministe

On sait que la solution d'une équation aux différences telle que

$$[(1 - B)^{r+d} z_t] - \mu = (1 - B)^r e_t \quad (7.7)$$

est de la forme

$$\nabla^d z_t - \sum_{j=0}^r \beta_j t^j = e_t$$

avec $\beta_r = \mu / (r!)$. On peut donc utilement inclure une tendance déterministe dans un modèle, par exemple lorsqu'un modèle ARIMA estimé sur base de données fournit une équation ressemblant à (7.7). Plus généralement, le modèle ARIMA avec tendance polynomiale additive s'écrit

$$\phi(B)[\nabla^d z_t - \mu_t] = \theta(B)e_t,$$

où μ_t est un polynôme en t de degré r . On peut utiliser une paramétrisation par le biais des polynômes orthogonaux, par exemple, pour un polynôme de degré 3:

$$\mu_t = \beta_0 + \beta_1(t - \nu) + \beta_2 \left[(t - \nu)^2 - \frac{n^2 - 1}{12} \right] + \beta_3 \left[(t - \nu)^3 - \frac{3n^2 - 7}{20}(t - \nu) \right],$$

où $\nu = (n + 1)/2$ est le milieu de l'intervalle des observations, après application des différences éventuelles.

Il est évident qu'une s.c. présentant une tendance polynomiale sera mieux représentée de cette façon que par un modèle avec différence régulière puisque celui-ci risque d'être non inversible, comme l'est (7.7).

7.7. La différence saisonnière et la composante saisonnière déterministe

La solution d'une équation aux différences telle que

$$(1 - B^s)z_t = (1 - B^s)e_t$$

n'est pas de la forme $z_t = e_t$ mais bien $z_t = \mu + m_t + e_t$, où m_t est une fonction périodique de période s , donc $m_t = m_{t-s}$ pour tout t , telle que $\sum_{t=1}^s m_t = 0$ et μ est une constante (Abraham et Box, 1978).

Plus généralement, on a le modèle ARIMA avec composante saisonnière déterministe

$$\phi(B)[\nabla^d z_t - m_t] = \theta(B)e_t.$$

La fonction m_t peut être paramétrée à l'aide de $(s-1)$ valeurs, par exemple en $t = 1, 2, \dots, s-1$, où à l'aide des $(s-1)$ coefficients de Fourier, c'est-à-dire les coefficients de la décomposition finie de Fourier

$$m_t = \sum_{j=1}^{\lfloor (s-1)/2 \rfloor} a_j \sin \frac{2\pi jt}{s} + \sum_{j=1}^{\lfloor s/2 \rfloor} b_j \cos \frac{2\pi jt}{s}, \quad (7.8)$$

où $\lfloor x \rfloor$ est le plus grand entier contenu dans x . Afin d'obtenir un modèle parcimonieux, il peut s'avérer utile de combiner partiellement les deux paramétrisations sous la forme

$$m_t = m'_t + m''_t,$$

où m'_t et m''_t sont deux fonctions périodiques de période s , dépendant respectivement de k' et k'' paramètres, $k' + k'' \leq s-1$ et définies comme suit : on choisit k' instants t_i dans $\{1, 2, \dots, s\}$. On introduit k' paramètres c_i ($i = 1, \dots, k'$) qui servent à définir m_t comme suit :

$$m'_t = \begin{cases} c_i, & \text{si } t = t_i \quad (i = 1, \dots, k') \\ -\frac{1}{s-k'} \sum c_i & \text{si } t \notin \{t_1, \dots, t_{k'}\} \end{cases}$$

de sorte que $\sum_{t=1}^s m'_t = 0$. On définit ensuite m''_t comme étant la somme de k'' termes de (7.8), diminuée d'une constante telle que $\sum_{t=1}^s m''_t = 0$.

Une justification de l'introduction de la composante saisonnière déterministe est donnée par Abraham et Box (1978). Elle convient mieux que la différence saisonnière lorsque le mouvement saisonnier est rigide. Au contraire, la différence saisonnière permet une meilleure représentation d'un mouvement saisonnier qui évolue au cours du temps. Dans certains cas il y a intérêt à combiner les deux procédés comme le montre l'exemple suivant. Supposons que $\{z_t - m_t(1 + n_t); t \in \mathbb{Z}\}$ soit un p.s.s., où m_t et n_t sont des fonctions périodiques de période s . Appliquons l'opérateur $\nabla_s = 1 - B^s$. Il vient $\nabla_s z_t - \nabla_s m_t - \nabla_s m_t n_t = \nabla_s z_t - s m_t n_t$. Posons $m_t^* = s m_t n_t$. On voit qu'on a intérêt à considérer $\nabla_s z_t - m_t^*$. Une s.c. engendrée par ce p.a. apparaîtra comme ayant un mouvement saisonnier évoluant de façon

régulière, linéairement de fait. Plus précisément, m_t^* est l'accroissement annuel du coefficient saisonnier du mois t .

7.8. L'estimation des paramètres: la méthode conditionnelle

Le modèle ARIMA généralisé contient un certain nombre de paramètres:

- la variance σ^2 des innovations normalisées e_t ,
- les paramètres inclus dans les polynômes AR et MA, y compris ceux qui caractérisent l'évolution en fonction du temps,
- les paramètres inclus dans les fonctions d'intervention y_t^I , y_t^D , y_t^M et y_t^S ,
- le paramètre λ de la transformation de puissance,
- les paramètres inclus dans les fonctions f_t et g_t ,
- les paramètres inclus dans la tendance déterministe μ_t ,
- les paramètres inclus dans le saisonnier déterministe m_t .

Nous symbolisons par le vecteur $\boldsymbol{\beta}$ l'ensemble de ces paramètres à l'exception de σ^2 . On dispose d'une s.c. (z_1, \dots, z_n) à partir de laquelle on veut estimer $\boldsymbol{\beta}$ et σ^2 au moyen du critère du maximum de vraisemblance, sous l'hypothèse que les e_t sont des v.a.s. indépendantes de moyenne 0 et de variance σ^2 . Notons $\ell = d + Ds$, où d , D et s sont respectivement l'ordre de la différence ordinaire, l'ordre de la différence saisonnière et la période saisonnière. Remarquons que z_t n'est pas nécessairement une v.a.s. lorsque $\ell > 0$. Dans ce qui suit, on considère donc la f.v. conditionnelle par rapport à $\mathbf{z}_0 = (z_1, \dots, z_\ell)^\top$. Dans une première approche, on développe la méthode du m.v.c., ce qui consiste à supposer fixées, non seulement z_1, \dots, z_ℓ , mais également les z_t et les e_t qui doivent servir de valeurs initiales à l'équation aux différences (7.2), c'est-à-dire $z_{\ell+1}, \dots, z_{\ell+p}$ et $e_{\ell+p-q+1}, \dots, e_{\ell+p}$. Notons \mathbf{v}_0 le vecteur de ces valeurs initiales. La f.v. s'écrit alors, en posant $h = \ell + p + 1$ et $N = n - (\ell + p)$

$$L(\boldsymbol{\beta}, \sigma^2; z | \mathbf{z}_0, \mathbf{v}_0) = (2\pi\sigma^2)^{-N/2} \exp \left[-\frac{1}{2\sigma^2} \sum_{t=h}^n e_t^2 \right].$$

On doit exprimer e_t en fonction des observations z_h, \dots, z_n , ce qui introduit en facteur le jacobien de la transformation $J(\boldsymbol{\beta})$, c'est-à-dire le déterminant de la matrice des $(\partial e_t / \partial z_u)$ pour $t, u = h, \dots, n$. Cette matrice est triangulaire inférieure puisque e_t ne dépend pas des z_u pour $u > t$. Le jacobien est donc le produit des éléments diagonaux

$$J(\boldsymbol{\beta}) = \prod_{t=h}^n \frac{\partial e_t}{\partial z_t}.$$

Or

$$\frac{\partial e_t}{\partial z_t} = [g_t y_t^S]^{-1} \frac{\partial a_t}{\partial z_t}.$$

$$\frac{\partial a_t}{\partial z_t} = \frac{\partial w_t}{\partial z_t} = f_t^{-1} \frac{\partial \tilde{C}_\lambda(z_t - y_t^I)}{\partial z_t} = f_t^{-1} (z_t - y_t^I + c)^{\lambda-1}$$

compte tenu de 7.5). Dès lors

$$J(\boldsymbol{\beta}) = \prod_{t=h}^n [g_t^{-1} f_t^{-1} (y_t^S)^{-1} (z_t - y_t^I + c)^{\lambda-1}].$$

Normalisons les fonctions f_t , g_t et y_t^S de façon telle que

$$\prod_{t=1}^n f_t = \prod_{t=1}^n g_t = \prod_{t=1}^n y_t^S = 1.$$

Par conséquent,

$$J(\boldsymbol{\beta}) = \prod_{t=1}^n (z_t - y_t^I + c)^{\lambda-1} = G^{N(\lambda-1)},$$

où G est la moyenne géométrique des $(z_t - y_t^I + c)$, pour $t = h, \dots, n$. On a donc

$$L(\boldsymbol{\beta}, \sigma^2; z|z_0, v_0) = (2\pi\sigma^2)^{-N/2} G^{N(\lambda-1)} \exp \left[-\frac{1}{2\sigma^2} \sum_{t=h}^n e_t^2(\boldsymbol{\beta}) \right],$$

où $e_t(\boldsymbol{\beta})$ est la valeur de e_t déterminée en fonction des z_t , à l'aide des relations de récurrence (7.1-3), pour les valeurs $\boldsymbol{\beta}$ des paramètres. Posons

$$\sigma_* = \sigma G^{-(\lambda-1)}$$

de sorte que

$$L(\boldsymbol{\beta}, \sigma_*^2; z|z_0, v_0) = (2\pi\sigma_*^2)^{-N/2} \exp \left[-\frac{1}{2\sigma_*^2} \sum_{t=h}^n [e_t^*(\boldsymbol{\beta}) G^{-(\lambda-1)}]^2 \right].$$

Notons que $e_t^*(\boldsymbol{\beta}) = e_t(\boldsymbol{\beta}) G^{-(\lambda-1)}$ s'obtient simplement en remplaçant dans les relations (7.1-3) $\tilde{C}_\lambda(z_t - y_t^I)$ par $C_\lambda(z_t - y_t^I)$, défini par (7.6). Le logarithme de la f.v. concentrée

$$\log L = \log L(\boldsymbol{\beta}, \sigma_*^2; z|z_0, v_0) = -\frac{N}{2} \log(2\pi\sigma_*^2) - \frac{S(\boldsymbol{\beta})}{2\sigma_*^2} \quad \text{avec} \quad S(\boldsymbol{\beta}) = \sum_{t=h}^n [e_t^*(\boldsymbol{\beta})]^2$$

peut être maximisé par rapport à σ_*^2 . Or

$$\frac{\partial \log L}{\partial \sigma_*^2} = -\frac{N}{2} \frac{1}{\sigma_*^2} + \frac{S(\boldsymbol{\beta})}{2\sigma_*^4}$$

s'annule pour $\sigma_*^2 = S(\boldsymbol{\beta})/N$. On procède donc en deux étapes :

(i) on minimise $S(\boldsymbol{\beta})$ par rapport à $\boldsymbol{\beta}$ et on note $\hat{\boldsymbol{\beta}}$ l'estimateur ainsi obtenu;

(ii) on détermine l'estimateur du m.v. de σ_* , $\hat{\sigma}_*^2 = S(\hat{\boldsymbol{\beta}})/N$.

L'étape (i) requiert l'utilisation d'un algorithme d'optimisation. Compte tenu de ce que la fonction à minimiser a la forme d'une somme de carrés, on peut songer à l'algorithme de Marquardt (1963) qui réalise un compromis entre la méthode de la plus grande pente et celle de Gauss-Newton, qui utilise une approximation quadratique de la fonction.

En plus de $\hat{\boldsymbol{\beta}}$ et de $\hat{\sigma}_*^2$, cet algorithme fournit une estimation de la matrice de covariance de $\hat{\boldsymbol{\beta}}$, calculée sur base de la théorie asymptotique des estimateurs du maximum de vraisemblance, en supposant que celle-ci soit justifiée, ce qui n'est pas évident lorsque le processus ARIMA est évolutif au lieu d'être constant. On part de

$$\begin{aligned} \|\hat{\boldsymbol{\beta}}\|^2 &= - \left[E \left(\frac{\partial^2 \log L}{\partial \boldsymbol{\beta}^\top \partial \boldsymbol{\beta}} \right) \right]^{-1} \\ &= 2\sigma_*^2 \left[E \left(\frac{\partial^2 S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^\top \partial \boldsymbol{\beta}} \right) \right]^{-1}. \end{aligned}$$

Or

$$\frac{\partial^2 S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^\top \partial \boldsymbol{\beta}} = 2 \sum_{t=h}^n \left[\frac{\partial e_t^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^\top} \frac{\partial e_t^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} + e_t^*(\boldsymbol{\beta}) \frac{\partial^2 e_t^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^\top \partial \boldsymbol{\beta}} \right]. \quad (7.9)$$

Dans le cas d'un processus ARIMA non généralisé, les paramètres de $\boldsymbol{\beta}$ sont employés exclusivement dans les coefficients de $e_{t-i}(\boldsymbol{\beta})$ et z_{t-i} pour $i \geq 1$, de sorte que $(\partial e_t^*(\boldsymbol{\beta})/\partial \boldsymbol{\beta}) = (\partial e_t(\boldsymbol{\beta})/\partial \boldsymbol{\beta})$ est une v.a. non corrélée avec $e_t(\boldsymbol{\beta})$. L'espérance mathématique du second terme de (7.9) est donc nulle, ce qui justifie l'estimation de $\|\hat{\boldsymbol{\beta}}\|^2$ par

$$\hat{\sigma}_*^2 \left[\sum_{t=h}^n \frac{\partial b_t^*(\hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}^\top} \frac{\partial b_t^*(\hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}} \right]^{-1}$$

comme fourni par l'algorithme de Marquardt. En revanche, dans le cas d'un modèle ARIMA généralisé, certains paramètres tels ceux contenus dans f_t , g_t , y_t^S ainsi que λ , apparaissent dans le coefficient de z_t . La simplification mentionnée plus haut ne survient donc plus, ce qui impose d'évaluer (7.9) en tenant compte de la dérivée partielle seconde (Brubacher, 1977).

7.9. L'estimation des paramètres: la méthode non conditionnelle

On veut déterminer $L(\boldsymbol{\beta}, \sigma^2; \mathbf{z}|\mathbf{z}_0)$ c'est-à-dire la f.v. conditionnelle à $\mathbf{z}_0 = (z_1, \dots, z_\ell)^\top$ mais sans spécifications de valeurs initiales. On suppose que la distribution liée de

$$x_t = \nabla^d \nabla_s^D \{ \tilde{C}_\lambda(z_t - y_t^I) / f_t \} \quad (t = \ell + 1, \dots, n)$$

... est normale, de vecteur moyenne \mathbf{M} et de matrice de covariance $\mathbf{\Gamma}$, de sorte que, en notant $\mathbf{x} = (x_{\ell+1}, \dots, x_n)$:

$$L(\boldsymbol{\beta}, \sigma^2; \mathbf{z} | \mathbf{z}_0) = (2\pi^2)^{-(n-\ell)/2} (\det \mathbf{\Gamma})^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mathbf{M})^\top \mathbf{\Gamma}^{-1} (\mathbf{x} - \mathbf{M}) \right\}.$$

Le problème est de déterminer \mathbf{M} et d'appliquer l'algorithme de la section 6.5 dans le cas où on se ramène à un processus ARMA constant ou l'algorithme de la section 6.3 dans le cas où on se ramène plutôt à un processus ARMA évolutif. On sera dans ce dernier cas dès le moment où $g_t \neq 1$ ou $y_t^S \neq 1$ pour au moins un t ou si au moins un des coefficients du sous-modèle ARMA dépend du temps. En revanche, la présence de $C_\lambda(z_t - y_t^f)$ ou de f_t n'a aucune incidence puisque ces composantes du modèle agissent au niveau du sous-modèle de la variable, donc avant que le sous-modèle ARMA n'entre en jeu.

Considérons donc la détermination de $E_0(x_t)$, l'espérance conditionnelle de x_t sachant \mathbf{z}_0 , ce qui est l'élément général de \mathbf{M} . De (7.1-3) on déduit les relations de récurrence:

$$\begin{aligned} E_0(x_t) &= E_0(w_t) + y_t^D + \mu_t + m_t, \\ E_0(w_t) &= E_0(a_t) + \sum_{i=1}^p \phi_{it} E_0(w_{t-i}) - \sum_{i=1}^q \theta_{it} E_0(a_{t-i}), \\ E_0(a_t) &= y_t^M g_t. \end{aligned}$$

Le problème est donc de trouver p valeurs initiales pour $E_0(w_t)$ ($t = \ell - p + 1, \dots, \ell$) qui permettront de déterminer les valeurs suivantes par récurrence. Si $E_0(a_t)$ est de forme connue pour tout t , on peut partir de la décomposition moyenne mobile infinie de $(w_t; t \in \mathbb{Z})$

$$w_t = \sum_{j=0}^{\infty} \psi_{jt} a_{t-j}.$$

En pratique, on utilisera une somme finie en négligeant les termes, tels que $|\psi_{jt}| < \varepsilon$ pour tout $j > J(\varepsilon)$, mettons, où ε est fixé à priori. On emploie alors

$$E_0(w_t) = \sum_{j=0}^{J(\varepsilon)} \psi_{jt} E_0(a_{t-j}).$$

Dans des cas particuliers, il est possible de simplifier ce calcul. Notons

$$\lambda_t = E_0(a_t) - \sum_{i=1}^q \theta_{it} E_0(a_{t-i}).$$

Si $y_t^M = 0$ pour $t \leq \ell$, alors $\lambda_t = 0$ et donc $E_0(w_t) = 0$ pour $t \leq \ell$. Si $y_t^M = \mu'$ est une constante non nulle pour $t \leq \ell$, et si l'on suppose comme à la section 6.4 que $\phi_{it} = \phi_{i\ell}$, $\theta_{it} = \theta_{i\ell}$, $g_t = g_\ell$ pour $t \leq \ell$, alors λ_t est une constante $\lambda = \lambda_\ell$, et donc

$$E_0(w_t) = \frac{\lambda}{1 - \sum_{i=0}^p \phi_{i\ell}} = \frac{1 - \sum_{j=0}^q \theta_{j\ell}}{1 - \sum_{i=0}^p \phi_{i\ell}} g_\ell \mu', \quad t \leq \ell. \quad (7.10)$$

Dans les mêmes conditions concernant les $\phi_{i\ell}$, les $\theta_{j\ell}$ et les g_t , on peut également considérer le cas où $y_t^M = m'_t + \mu'$, où m'_t est une fonction périodique, de période s et dont la somme sur une période vaut 0 et μ' est une constante non nulle. Notons $\phi_\ell(B)$ et $\theta_\ell(B)$ les polynômes AR et MA, respectivement, à l'instant $t = \ell$ et $\psi_\ell(B) = \phi_\ell^{-1}(B)\theta_\ell(B) = \sum_{j=0}^{\infty} \psi_{j\ell}B^j$. Généralisant (7.10), il vient

$$\begin{aligned} E_0(w_t) &= g_\ell \psi_\ell(B)(m'_t + \mu') \\ &= g_\ell \left[\sum_{j=0}^{s-1} \left(\sum_{i=0}^{\infty} \psi_{i\ell} \right) m'_{t-h} + \phi^{-1}(1)\theta(1)\mu' \right]. \end{aligned}$$

Bibliographie

1. ABDREBBO, N.A. and PRIESTLEY, M.B. (1967), On the prediction of non-stationary processes, *J. Roy. Statist. Soc. Ser. B* 29, 570-585.
2. ABRAHAM, B. and BOX, G.E.P. (1978), Deterministic and forecast-adaptative time-dependent models, *J. Roy. Statist. Soc. Ser. C, Appl. Statist.* 27, 120-130.
3. AKAIKE, H. (1974), A new look at the statistical model identification, *IEEE Trans. Autom. Control* AC-19, 716-723.
4. ANDERSON, B.D.O. and MOORE, J.B. (1979), *Optimal Filtering*, Prentice-Hall, Englewood Cliffs.
5. ANDERSON, B.D.O., MOORE, J.B. and LOO, S.G. (1969), Spectral factorization of time-varying covariance functions, *IEEE Trans. Information Theory*, IT-15, 550-557.
6. ANDERSON, O.D. (1975), The recursive nature of the stationarity and invertibility restraints on the parameters of mixed autoregressive-moving average processes, *Biometrika* 62, 704-706.
7. ANDERSON, O.D. (1977), Time series analysis and forecasting: a further look at the Box-Jenkins approach, *Cahiers du Centre d'études de recherche opérationnelle* 19, 223-256.
8. ANDERSON, O.D. (1979a), The autocovariance structures associated with general unit circle nonstationary factors in the autoregressive operators of otherwise stationary ARMA time series models, *Cahiers du Centre d'études de recherche opérationnelle* 21, 221-237.
9. ANDERSON, O.D. (1979b), Some sample autocovariance function results for a once integrated q th-order moving average process, *Statistica* 39, 287-299.
10. ANDERSON, T.W. (1971), *The Statistical Analysis of Time Series*, Wiley, New York.
11. ANSLEY, C.F. (1979), An algorithm for the exact likelihood of a mixed autoregressive-moving average process, *Biometrika* 66, 59-65.
12. ANSLEY, C.F. (1980), Computation of the theoretical autocovariance function for a vector ARMA process, *J. Statist. Comput. Simul.* 12, 15-24.
13. ANSLEY, C.F. and KOHN, R. (1983), Exact likelihood of vector autoregressive-moving average process with missing or aggregated data, *Biometrika* 70, 275-278.
14. ANSLEY, C.F. and NEWBOLD, P. (1980), Finite sample properties of estimators for autoregressive moving average models, *J. Econometrics* 13, 159-183.
15. ANSLEY, C.F., SPIVEY, W.A. and WROBLESKI, W.J. (1977), A class of transformations for Box-Jenkins seasonal models, *J. Roy. Statist. Soc. Ser. C, Appl. Statist.* 26, 173-178.
16. ÅSTRÖM, K.J. (1970), *Introduction to Stochastic Control Theory*, Academic Press, New York.
17. BACHELET, D. et MORLAT, G. (1966), Modèle à deux aléas pour des chroniques économiques, *Revue française de recherche opérationnelle* 40, 275-289.
18. BAGSHAW, M. and JOHNSON, R.A. (1977), Sequential procedures for detecting parameter changes in a time series model, *J. Amer. Statist. Assoc.* 72, 593-597.

19. BEGUIN, J.-M., GOURIEROUX, C. et MONFORT, A. (1980), Identification of a mixed autoregressive-moving average process: the corner method, In O.D. Anderson (ed): *Time Series*, North-Holland, Amsterdam, 423-436.
20. BELL, William (1983), A computer program for detecting parameter changes in time series, *1983 Proceedings of the Business and Economic Statistics Section*, American Statistical Association, Washington, 634-649.
21. BHATTACHARYYA, M.N. and LAYTON, A.P. (1979), Effectiveness of seat belt legislation on the Queensland road toll - An Australian case study in intervention analysis, *J. Amer. Stat. Assoc.* 74, 596-603.
22. BORDIGNON, S. et MASAROTTO, G. (1983), Una classe di modelli non stationari, *Statistica* 43, 83-103.
23. BOX, G.E.P. and COX, D.R. (1964), An analysis of transformations, *J. Roy. Statist. Soc. Ser. B* 26, 211-243.
24. BOX, G.E.P. and JENKINS, G.M. (1976), *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco (revised edition).
25. BOX, G.E.P. and PIERCE, D.A. (1970), Distribution of residuals autocorrelations in autoregressive-integrated moving average time series models, *J. Amer. Statist. Assoc.* 65, 1509-1526.
26. BOX, G.E.P. and TIAO, G.C. (1975), Intervention analysis with applications to economic and environmental problems, *J. Amer. Statist. Assoc.* 70, 70-79.
27. BROWN, R.G. (1959), *Statistical Forecasting for Inventory Control*, McGraw-Hill, New York.
28. BRUBACHER, S.R. (1977), Applications and extensions of the Box-Jenkins model, Communication at the Public Utilities Forecasting Conference, Brownson-on-Windermere.
29. CHANG, I. and TIAO, G.C. (1983), Effect of exogenous interventions on the estimation of time-series parameters, *1983 Proceedings of the Business and Economic Section*, American Statistical Association, Washington, 532-537.
30. CHARLES, B. et ROY, R. (1980), Calcul fonctionnel et séries chronologiques, *Statistique et analyse des données* 1, 33-49.
31. CRAMÉR, H. (1961), On some classes of non-stationary stochastic processes, *Proceedings of the fourth Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley and Los Angeles, Vol. 2, 57-78.
32. de FALGUEROLLES, A. (1979), Autorégression avec écrêtement, *Rev. Statist. Appl.* 27, 29-44.
33. de FALGUEROLLES, A. (1980), Identification et estimation préliminaire des modèles ARMA, *Cahiers du Centre d'études de recherche opérationnelle* 22, 149-158.
34. DE SCHUTTER, A. (1977), Une généralisation des concepts spectraux non stationnaires, *Cahiers du Centre d'études de recherche opérationnelle* 19, 365-377.
35. DOOB, J. L. (1983), *Stochastic Processes*, Wiley, New York.
36. DUNSMUIR, W. (1981), Estimation of periodically varying means and standard deviations in time series data, *J. Time Series Anal.* 2, 129-153.
37. DURBIN, J. (1960), The fitting of time series models, *Rev. Inst. Int. Stat.* 28, 233-244.
38. FOX, A.J. (1972), Outliers in time series, *J. Roy. Statist. Soc. Ser. B* 34, 350-363.
39. FULLER, W.A. (1976), *Introduction to Statistical Time Series*, Wiley, New York.
40. GARDNER, G., HARVEY, A.C. and PHILLIPS, G.D.A. (1980), Algorithm AS 154, Algorithm for exact maximum likelihood estimation of autoregressive-moving average models by means of Kalman filtering, *J. Roy. Statist. Soc. Ser. C, Appl. Statist.* 29, 311-322.
41. GOODMAN, M. L. (1974), A new look at higher-order exponential smoothing for forecasting, *Operations Research* 22, 880-888.
42. GOODWIN, G.C. and PAYNE, R. L. (1977), *Dynamic System Identification: Experiment Design and Data Analysis*, Academic Press, New York.
43. GOURIEROUX, Ch. et MONFORT, A. (1983), *Cours de séries temporelles*, Economica, Paris.
44. GRANGER, C.W.J. and NEWBOLD, P. (1977), *Forecasting Economic Time Series*, Academic Press, New York.

45. GRENANDER, U. and ROSENBLATT, M. (1956), *Statistical Analysis of Stationary Time Series*, Almqvist and Wiksell, Stockholm.
46. HALLIN, M. (1978), Mixed autoregressive-moving average multivariate processes with time-dependent coefficients, *J. Multivariate Anal.* 8, 567-572.
47. HALLIN, M. (1984), Spectral factorization of nonstationary moving average processes, *Ann. Statist.* 12, 172-192.
48. HALLIN, M. (1986), Nonstationary q -dependent processes and time-varying moving average models. Invertibility properties and the forecasting problem, *Advances in Applied Probability* 18, 170-210.
49. HALLIN, M. et INGENBLEEK, J.-F. (1982), The mode-building problem for nonstationary multivariate autoregressive processes, In O. D. Anderson (ed.): *Time Series Analysis: Theory and Practice I*, North-Holland, Amsterdam, 599-606.
50. HALLIN, M. and INGENBLEEK, J.-F. (1983), Nonstationary Yule-Walker equations, *Statistics and Probability Letters* 1, 189-195.
51. HALLIN, M. et MÉLARD, G. (1977), Indéterminabilité pure et inversibilité des processus autorégressifs - moyenne mobile à coefficients dépendant du temps, *Cahier du Centre d'étude de recherche opérationnelle* 19, 385-392.
52. HAMDI, A. (1982), Identification des processus ARMA à coefficients dépendant du temps, Thèse de 3e cycle, Université Paul Sabatier, Toulouse.
53. HANNAN, E. J. (1969), The identification of vector mixed autoregressive-moving average systems, *Biometrika* 56, 223-225.
54. HANNAN, E.J. (1970), *Multiple Time Series*, Wiley, New York.
55. HARRISON, P.J. (1967), Exponential smoothing and short-term sales forecasting, *Management Science* 13, 821-842.
56. HARRISON, P.J. and STEVENS, C.F. (1976), Bayesian forecasting, *J. Roy. Statist. Soc. Ser. B* 38, 205-247.
57. HARVEY, A.C. and PHILLIPS, G.D.A. (1979), Maximum likelihood estimation of regression models with autoregressive-moving average disturbances, *Biometrika* 66, 49-58.
58. HERBST, L.J. (1963), A test for variance heterogeneity in the residuals of a Gaussian moving average, *J. Roy. Statist. Soc. Ser. B* 25, 451-454.
59. HILL, B.M. (1963), The three-parameter lognormal distribution and Bayesian analysis of a point-source epidemic, *J. Amer. Statist. Assoc.* 58, 72-84.
60. HILLMER, S.C. and TIAO, G.C. (1979), Likelihood function of stationary multiple autoregressive moving average models, *J. Amer. Statist. Assoc.* 74, 652-661.
61. HILLMER, S.C. and TIAO, G.C. (1982), An ARIMA-model-based approach to seasonal adjustment, *J. Amer. Statist. Assoc.* 77, 63-70.
62. HOLT, C.C., MODIGLIANI, F., MUTH, J.F. and SIMON, H.A. (1963), *Planning Production, Inventories and Work Force*, Prentice-Hall, New Jersey.
63. HUYBERECHTS, S. (1983), Interprétation des fonctions matricielles de régression partielle dans les modèles ARMA multivariés, *Cahiers du Centre d'étude de recherche opérationnelle* 25, 3-8.
64. JAZWINSKI, A.H. (1970), *Stochastic Processes and Filtering Theory*, Academic Press, New York.
65. KALMAN, R.E. (1960), A new approach to linear filtering and prediction problems, *Jour. of Basic Eng. (Trans A.S.M.E.) Series D* 82, 35-95.
66. KARHUNEN, K. (1947), Über lineare Methoden in den Wahrscheinlichkeitsrechnung, *Ann. Acad. Scient. Fennicae, Ser. A-I* 37, 7-79.
67. KASHYAP, R.L. and RAO, A.R. (1976), *Dynamic Stochastic Models from Empirical Data*, Academic Press, New York.
68. KESTEMONT, R.M. (1983), Sur l'évaluation de la fonction de vraisemblance du modèle ARMA stationnaire, Rapport No. 33, Institut de mathématique pure et appliquée, Université Catholique de Louvain.
69. KHABIE-ZEITOUNE, E. (1980), Closed form likelihood in multivariate weakly stationary time series, In O.D. Anderson (ed.) : *Time Series*, North-Holland, Amsterdam, 311-318.

70. KIEHM, J.-L., LEFÈVRE, Cl. et MÉLARD, G. (1980), Forme rétrospective d'un modèle FARIMAG, *Cahiers du Centre d'études de recherche opérationnelle* 22, 375-383.
71. KOHN, R. and ANSLEY, C.F. (1982), A note on obtaining the theoretical autocovariance of an ARMA process, *J. Statist. Comput. Simul.* 15, 273-283.
72. KOZIN, F. and NAKAJIMA, F. (1980), The order determination problem for linear time-varying AR models, *IEEE Trans. Autom. Control*, AC-25, 250-257.
73. LEVINSON, N. (1949), A heuristic exposition of Wiener's mathematical theory of prediction and filtering, In N. Wiener: *Extrapolation, Interpolation and Smoothing of Stationary Time Series, with Engineering Applications*, Wiley, New York.
74. LIBERT, G. (1977), Comparaison et interprétation de deux méthodes de traitement de séries chronologiques, 10th European Meeting of Statisticians, August 10-12, Louvain.
75. LINDBERGER, N.A. (1973), Comments on "On estimating the orders of an autoregressive moving-average process with uncertain observations", *IEEE Trans. Autom. Control*, AC-18, 689-691.
76. LINDQUIST, A. (1974), A new algorithm for optimal filtering of discrete-time stationary processes, *Siam J. Contr.* 12, 736-746.
77. LJUNG, G. and BOX, G.E.P. (1978), On a measure of lack of fit in time series models, *Biometrika* 65, 297-303.
78. LJUNG, G.M. and BOX, G.E.P. (1979), The likelihood function of stationary autoregressive-moving average models, *Biometrika* 66, 265-270.
79. LOËVE, M. (1963), *Probability Theory*, Van Nostrand, Princeton, N.J. (3rd edition).
80. McKENZIE, E. (1974), A comparison of standard forecasting systems with the Box-Jenkins approach, *The Statistician* 23, 107-116.
81. McKENZIE, E. (1976), A comparison of some standard seasonal forecasting systems, *The Statistician* 25, 3-14.
82. McLEOD, A.I. (1975), Derivation of the theoretical autocovariance function of autoregressive-moving average time series, *J. Roy. Statist. Soc. Ser. C, Appl. Statist.* 24, 255-256.
83. MARQUARDT, D.W. (1963), An algorithm for least-squares estimation of non-linear parameters, *Journal of the Society of Industrial Applied Mathematics* 11, 431-441.
84. MÉLARD, G. (1975), Processus purement indéterminables à paramètre discret - approches fréquentielle et temporelle, Dissertation doctorale, Faculté des sciences, Université Libre de Bruxelles.
85. MELARD, G. (1977), Sur une classe de modèles ARIMA dépendant du temps, *Cahiers du Centre d'études de recherche opérationnelle* 19, 285-295.
86. MÉLARD, G. (1981), On an alternative model for intervention analysis, In O.D. Anderson and M.R. Perryman (eds.): *Time Series Analysis*, North-Holland, Amsterdam, 345-354.
87. MÉLARD, G. (1982a), The likelihood function of a time-dependent ARMA model, in O.D. Anderson and M.R. Perryman (eds.): *Applied Time Series Analysis, Proceedings of the International conference held at Houston, Texas, August 1981*, North-Holland, Amsterdam, 229-239.
88. MÉLARD, G. (1982b), Software for time series analysis, In H. Caussinus, P. Ettinger et R. Tomassone (eds.): *Compstat 1982, Part I, Proceedings in Computational Statistics*, Physica-Verlag, Vienna, 336-341.
89. MÉLARD, G. (1982c), L'application de la méthode du maximum de vraisemblance dans des modèles de séries chronologiques, *Cahiers du Centre d'études de recherche opérationnelle* 24, 353-365.
90. MÉLARD, G. (1983), ANSECH - Logiciel pour l'analyse des séries chronologiques, Notes de cours, Presses universitaires de Bruxelles, (5e édition).
91. MÉLARD, G. (1984), Algorithm AS197: A fast algorithm for the exact likelihood of autoregressive moving average models, *J. Roy. Statist. Soc. Ser. C, Appl. Statist.* 33, 104-114.
92. MÉLARD, G. (1985), Illustration of the use of a general time series model, In O.D. Anderson (ed.): *Time Series Analysis: Theory and Practice 7*, Proceedings of the International Conference held at Toronto, August 1983, North-Holland, Amsterdam, 1985, 53-75.

93. MÉLARD, G. and KIEHM, J.-L. (1981), ARIMA models with time-dependent coefficients for economic time series, In O.D. Anderson and M.R. Perryman (eds.) : *Time Series Analysis*, North-Holland, Amsterdam, 355-363.
94. MÉLARD, G. et ROULAND, O. (1986), Sélection d'une méthode de prévision par l'emploi du modèle ARIMA sous-jacent, *R.A.I.R.O Recherche opérationnelle/Operations Research* 20, 89-113.
95. MILLER, K.S. (1968), *Linear Difference Equations*, Benjamin, New York.
96. MILLER, K.S. (1969), Nonstationary autoregressive processes, *IEEE Trans. Inform. Theory* IT-15, 315-316.
97. MORF, M., SIDHU, G.S. and KAILATH, T. (1974), Some new algorithms for recursive estimation in constant, linear, discrete-time systems, *IEEE Trans. Autom. Control* AC-19, 315-323.
98. NEWBOLD, P. (1974), The exact likelihood function for a mixed autoregressive-moving average process, *Biometrika* 61, 423-426.
99. NEWBOLD, P. (1981), Some recent developments in time series analysis, *Intern. Statist. Rev.* 49, 53-66.
100. NICHOLLS, D.F. and HALL, A.D. (1979), The exact likelihood function of multivariate autoregressive-moving average models, *Biometrika* 66, 259-264.
101. OZAKI, T. (1977), On the order determination of ARIMA models, *J. Roy. Statist. Soc. Ser. C, Appl. Statist.* 26, 290-301.
102. PAGANO, M. (1978), On periodic and multiple autoregressions, *Ann. Statist.* 6, 1310-1317.
103. PARZEN, E. (1967), Statistical inference on time series by Hilbert space methods I, In E. Parzen: *Time Series Analysis Papers*, Wiley, New York, 251-382.
104. PEARLMAN, J.G. (1980), An algorithm for the exact likelihood of a high-order autoregressive-moving average process, *Biometrika* 67, 232-233.
105. PRIESTLEY, M.B. (1980), State-dependent models: a general approach to non-linear time series analysis, *Cahiers du Centre d'études de recherche opérationnelle* 22, 285-307 and *J. Time Series Anal.* 1, 47-71.
106. PRIESTLEY, M.B. (1981), *Spectral Analysis and Time Series* (2 volumes), Academic Press, New York.
107. QUENOUILLE, M.H. (1957), *The Analysis of Multiple Time Series*, Griffin, London.
108. RAVEH, A. (1980), Least potytone analysis (LPA): a nonmetric technique for analysing time series data, In O.D. Anderson (ed.) : *Time Series*, North Holland, Amsterdam, 335-356.
109. RISSANEN, J. (1973), Algorithms for triangular decomposition of block Hankel and Toeplitz matrices with application to factoring positive matrix polynomials, *Math. Comput.* 27, 147-154.
110. ROBINSON, E.A. (1967), *Multi-channel Time Series Analysis*, Holden-Day, San Francisco.
111. ROY, R. (1980), Analyse et prévision de séries chronologiques, Notes du cours IFT 6631, Département d'informatique et de recherche opérationnelle, Université de Montréal.
112. ROY, R. (1982), On a stepwise method for detecting nonstationarity, Technical Report # 451, Département d'informatique et de recherche opérationnelle, Université de Montréal.
113. ROY, R. et PELLERIN, J. (1982), On long term air quality trends and intervention analysis, *Atmosphere Environment* 16, 161-169.
114. ROZANOV, Y.A. (1967), *Stationary Random Processes*, Holden Day, San Francisco.
115. RUDIN, W. (1966), *Real and Complex Analysis*, McGraw Hill, New York.
116. STOER, J. and BULIRSCH, R. (1980), *Introduction to Numerical Analysis*, Springer Verlag, New York.
117. SUBBA RAO, T. (1970), The fitting of non-stationary time-series models with time-dependent parameters, *J. Roy. Statist. Soc. Ser. B* 32, 312-322.
118. TAYLOR, A.E. (1958), *Introduction to Functional Analysis*, Wiley, New York.
119. TIAO, G.C. and BOX, G.E.P. (1981), Modeling multiple time series with applications, *J. Amer. Statist. Assoc.* 76, 802-816.
120. TYSSDAL, J.S. and TJØSTHEIM, D. (1982), Autoregressive processes with a time dependent variance, *J. Time Series Anal.* 3, 209-217.

121. VAN WINKEL, E.G.F. (1976), Some more observations in adaptative forecasting, In M. Roubens (ed.): *Proceedings of the Second European Congress on Operations Research*, November 29 - December 1, 1976, Stockholm, North-Holland, Amsterdam.
122. WEGMAN, E.J. (1974), Some results on non stationary first order autoregression, *Technometrics* 16, 321-322.
123. WHITTLE, P. (1965), Recursive relations for predictors of non-stationary processes, *J. Roy. Statist. Soc. Ser. B* 27, 523-532.
124. WILSON, G.T. (1979), Some efficient computational procedures for high order ARMA models, *J. Statist. Comput. Simul.* 8, 303-309.
125. WINTERS, P.R. (1960), Forecasting sales by exponentially. weighted moving averages, *Management Sci.* 6, 324-342.
126. WOLD, H. (1954), *A Study in the Analysis of Stationary Time Series*, Almqvist and Wicksell, Uppsala (2nd edition).
127. ZELLNER, A. and PALM, F. (1974), Time series analysis and simultaneous equation econometric models, *J. Econometrics* 2, 17-54.

Glossaire

\mathbb{Z}	Ensemble des entiers	1
\mathbb{R}	Ensemble des réels	1
z	Variable étudiée (scalaire)	1
\mathbf{z}	Variable étudiée (vectorielle)	1
k	Nombre de dimensions de la série chronologique	1
s.c.	Série chronologique	1
p.a.	Processus aléatoire	1
n	Longueur de la série chronologique	1
f.d.	Fonction de distribution	3
$F(\cdot)$	Fonction de distribution	3
N	Loi normale	4
\perp	Non corrélation	4
$\perp\!\!\!\perp$	Indépendance	4
$E(\cdot)$	Espérance mathématique	7
v.a.s.	Variable aléatoire centrée du second ordre	8
var	Variance	8
cov	Covariance	8
$\ \cdot\ $	Norme d'un vecteur	8
l.i.m.	Limite en moyenne	9
$\langle \cdot, \cdot \rangle$	Produit intérieur, gramien	8
$\ \cdot\ ^2$	Carré de la norme	8
$\ \cdot\ ^2$	$= \langle \cdot, \cdot \rangle$	8
$P(\cdot, \cdot)$	Projection orthogonale	9
\oplus	Somme directe d'espaces	9
$(\cdot, \cdot; t \in \mathbb{Z})$	Processus aléatoire	11
\mathbb{C}	Ensemble des complexes	7
\top	Symbole de transposition	7
$ \cdot $	Valeur absolue ou module, selon le cas	8
$\bar{\cdot}$	Symbole de conjugaison	8
V_2	Espace de Hilbert des variables aléatoires du second ordre	8
\forall	Quel que soit	9
ssi	Si et seulement si	7
$\geq \mathbf{0}$	Matrice définie positive	10
$> \mathbf{0}$	Matrice définie strictement positive	10

\exists	Il existe	9
\in	Appartenant à	9
b.b.	Bruit blanc	12
p.a.s.	Processus aléatoire du second ordre	12
b.b.s.	Bruit blanc du second ordre	12
b.b.s.c.	Bruit blanc du second ordre constant	12
b.b.s.n.	Bruit blanc du second ordre normalisé	12
b.b.c.	Bruit blanc constant	12
b.b.n.	Bruit blanc normalisé	12
$V_2(\mathbf{z}; t)$	Sous-espace linéaire fermé sous-tendu par $(\mathbf{z}_s)_i; i = 1, \dots, k, s \leq t$	13
\subseteq, \subset	Inclusion	13
\cap	Intersection	13
V_t	$V_2(\mathbf{z}; t)$	13
p.p.i.	Processus purement indéterminable	13
p.d.	Processus déterminable	13
ARMA	Autorégressif moyenne mobile	18
AR	Autorégressif	18
MA	Moyenne mobile	18
$(\mathbf{a}_t; t \in \mathbb{Z})$	Processus innovation	18
ARMA(p, q)	ARMA d'ordre (p, q)	18
AR(p)	AR d'ordre p	18
MA(q)	MA d'ordre q	18
B	Opérateur de retard à droite	18
o.p.r.d.	Opérateur polynomial de retard à droite	18
∇	Opérateur de différence ordinaire	19
∇_s	Opérateur de différence saisonnière	19
\mathbf{I}_k	Matrice unité d'ordre k	18
FARMAG		25
FARIMAG		25
$\phi^{-1}(B)$	Inverse de l'o.p.r.d. constant $\phi(B)$	25
m.l.d.	Modèle linéaire dynamique	30
$\hat{\mathbf{z}}_t(h)$	Prévision d'horizon h fait en t	26
p.s.s.	Processus stationnaire du second ordre	37
MCC	Méthode des moindres carrés conditionnelle	62
MCN	Méthode des moindres carrés non conditionnelle	63
MV	Méthode du maximum de vraisemblance	63
s.c.c.	Somme des carrés conditionnelle	67
f.v.c.	Fonction de vraisemblance conditionnelle	68
s.c.n.	Somme des carrés non conditionnelle	67
f.v.	Fonction de vraisemblance	68
$(\Gamma_{ts}; t, s \in \mathbb{Z})$	Fonction de covariance	12
$(\gamma_h; h \in \mathbb{Z})$	Fonction d'autocovariance	33
$\phi_t(B)$	Opérateur AR (cas évolutif)	18
$\theta_t(B)$	Opérateur MA (cas évolutif)	18
$\phi_p(B)$	Opérateur AR de degré p (cas constant)	38
$\theta_q(B)$	Opérateur MA de degré q (cas constant)	38
$\mu_r^*(B)$	$B^r \mu_r(B^{-1})$	51
D	Disque unite	52
ARIMA		57
ρ_h	Autocorrélation de retard h	61