

N° d'ordre :

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITE DES SCIENCES ET DE LA TECHNOLOGIE HOUARI BOUMEDIENE

Faculté des Sciences - Mathématiques

Thèse

Magister en Mathématiques

Option

Recherche opérationnelle (méthodes stochastiques)

Présenté par

Aknouche abdelhakim

T H E M E

**ESTIMATION EN-LIGNE ET HORS-LIGNE
DES MODELES ARMA
PERIODIQUES**

Soutenue le : 11/07/2001 devant le jury composé de :

Monsieur A. Aïssani,	Professeur,	U. S. T. H. B. : Président
Monsieur M. Bentarzi,	Professeur,	U. S. T. H. B. : Directeur de Thèse
Monsieur K. Boukhetala,	Maître de Conférences,	U. S. T. H. B. : Examineur
Monsieur O. Anes,	Maître de Conférences,	I. N. P. S. : Examineur
Monsieur M. S. Maamra,	Chargé de cours,	U. S. T. H. B. : Examineur

Remerciements

Tout d'abord, je remercie **Dieu** pour son aide et son soutien.

En témoignage de reconnaissance et de gratitude, je voudrais exprimer mes remerciements les plus sincères et les plus vifs à monsieur **Mohamed Bentarzi**, professeur et ex. Directeur de l'institut de mathématiques de l'U.S.T.H.B., de m'avoir initié à la recherche, de m'avoir proposer ce sujet, que j'ai trouvé très intéressant, et de m'avoir soutenu et supporté tout au long de ce travail. Ses conseils, sa patience, l'assistance éclairée qu'il m'a prêté et sa disponibilité m'ont été d'une aide précieuse. Qu'il trouve ici ma profonde admiration.

Je voudrais également exprimer mes remerciements et ma reconnaissance à monsieur le professeur **A. Aïssani**, Doyen de la faculté des sciences à l'U.S.T.H.B., de m'avoir honoré par son aimable acceptation à présider le jury de cette thèse.

Je tiens aussi à remercier chaleureusement monsieur **K. Boukhetala**, Maître de conférences à l'U.S.T.H.B. de m'avoir fait l'honneur de figurer dans le jury de cette thèse, et de l'intérêt qu'il a porté à ce travail.

Je voudrais également témoigner ma grande reconnaissance et mes plus vifs remerciements à monsieur **O. Anes**, Maître de conférences à l'I.N.P.S., de m'avoir fait l'honneur par son acceptation d'appartenir au jury de cette thèse.

Je remercie, particulièrement monsieur **M.S. Maamra**, Chargé de cours à l'U.S.T.H.B., pour l'intérêt apporté à mon travail, pour ses critiques, son aide et pour son acceptation d'être au sein du jury de cette thèse.

Je saisis cette opportunité pour exprimer ma grande reconnaissance et ma profonde gratitude à toutes les personnes qui m'ont aidé, de loin ou de près, à élaborer ce travail, je citerai en particulier : **Zahir Hamoutène, Rebiha Ghezali, Fayçal Boukerdenna, Aboubakr Aknouche, Riad Hakkou, Abdenour Bouhouche, Toufik Zioui, Madjid Benhammouda, Sofiane Hamdache, Fayçal Jakki, Brahim Zahaf, 'Djilali Idir', Amer Zahaf, Maamri Massinissa, Chaaban Saïd, Aïmad Saadoun, Amel Bertouche, Sadek Léon, Hayet et Nadia Arbaoui.**

Enfin, il ne me paraît pas possible de citer ici nommément tous ceux avec qui j'ai eu contact durant la préparation de cette thèse, mais qu'ils ne seront jamais oubliés et qu'ils en soient aussi chaleureusement remerciés.

Abdelhakim Aknouche

Sommaire

Introduction générale

1. Introduction	1
2. Présentation et apport de la thèse	6

Première Partie Estimation hors-ligne

Chapitre 1 Estimation hors-ligne des modèles ARMA classiques

1. Introduction	9
2. Généralités	10
3. Estimation par la méthode des moments	10
3.1. Les équations de Yule-Walker.....	11
3.2. Estimateurs de Yule-Walker.....	11
4. Algorithmes récursifs pour l'estimation des moments	12
4.1. Algorithme de Durbin-Levinson pour l'estimation des paramètres d'un modèle AR.....	12
4.2. Algorithme de Wilson pour l'estimation des autocovariances d'un modèle ARMA	13
4.2.1. Cas autorégressif pur AR.....	13
4.2.2. Cas autorégressif moyenne mobile ARMA	14
5. Méthodes alternatives d'estimation dans les modèles ARMA.....	16
5.1. Critères	16
5.2. Evaluation des fonctions critères	16
5.2.1. Evaluation de la somme des carrés conditionnelle.....	17
5.2.2. Evaluation de la somme des carrés non conditionnelle.....	18
5.2.3. Evaluation de la fonction de vraisemblance conditionnelle.....	18

5.2.4.	Evaluation de la fonction de vraisemblance non conditionnelle.....	19
5.3.	L'optimisation	20
5.3.1.	Algorithme du gradient	21
5.3.2.	Algorithme de Gauss-Newton	22
5.3.3.	Algorithme de compromis de Marquardt	23
6.	Propriétés statistiques des estimateurs.....	24
5.1.	Matrice d'information de Fisher.....	24
5.2.	Propriétés statistiques de l'estimateur du MV.....	25

Chapitre 2 Processus périodiquement corrélés et Modèles ARMA périodiques

1.	Introduction	26
2.	Processus périodiquement corrélés	28
2.1.	Fonction d'autocovariance d'un processus périodiquement corrélé	29
2.2.	Fonction d'autocovariance empirique d'un processus périodiquement corrélé ...	29
2.3.	Processus multivarié stationnaire associé à un processus univarié périodiquement corrélé.....	30
2.4.	Relation entre les autocovariances d'un processus multivarié stationnaire et celles d'un processus périodiquement corrélé	30
2.5.	Processus bruit blanc périodique	31
3.	Prédiction des processus périodiquement corrélés.....	31
3.1.	Equation de prédiction	31
3.1.1.	Equation de prédiction prospective	32
3.1.2.	Equation de prédiction rétrospective	32
4.	Modèles ARMA périodiques	34
4.1.	Représentation Linéaire des processus périodiquement corrélés	34
4.2.	Représentation de Wold-Cramér	34
4.3.	Processus purement indéterminable et processus déterminable du second ordre..	35
4.4.	Modèle ARMA périodique (PARMA).....	36
4.5.	Modèle autorégressif périodique (PAR)	37
4.5.1	Equations de Yule-Walker périodiques	37
4.5.2.	Matrice d'information de Fisher limite d'un modèle PAR	38
4.6.	Modèles moyenne mobile périodique (PMA)	38

5. Méthodes d'études des propriétés théoriques des modèles ARMA périodiques.....	39
5.1. Technique <i>period-span lumping</i> et étude des modèles univariés ARMA périodiques	39
5.2. Modèle ARMA multivarié associé à un modèle ARMA univarié	40
5.3. Condition de causalité d'un modèle ARMA multivarié.....	41
5.4. Condition d'inversibilité d'un modèle ARMA multivarié.....	41
5.5. Condition de causalité d'un modèle ARMA périodique univarié.....	42
5.6. Condition d'inversibilité d'un modèle ARMA périodique univarié	42
5.7. Technique « order-span lumping » et étude des modèles univariés ARMA périodiques	42
5.8. Modèle multivarié moyenne mobile périodique d'ordre 1 associé à un modèle multivarié moyenne mobile périodique d'ordre q	43
5.9. Modèle multivarié autorégressif périodique d'ordre 1 associé à un modèle autorégressif périodique d'ordre p	45

Chapitre 3 Estimation hors-ligne des modèles ARMA périodiques

1. Introduction	47
2. Méthodes basées sur des transformations des données dans les modèles ARMA périodiques (méthodes des moments)	48
2.1. Méthodes basées sur les équations de Yule-Walker périodiques	49
2.1.1. Cas de modèles autorégressifs périodiques purs.....	49
2.1.2. Cas de modèles autorégressifs moyenne mobile périodiques	50
2.1.3. Exemple d'application.....	50
2.2. Estimation des moments par des méthodes récursives	51
3. Estimations dans les modèles PARMA par des méthodes utilisant directement les données	56
3.1. Critères	56
3.2. Evaluation des fonctions critères	56
3.2.1. Somme des carrés conditionnelle	57
3.2.2. Somme des carrés non conditionnelle	58
3.2.3. Fonction de vraisemblance conditionnelle	59
3.2.4. Fonction de vraisemblance non conditionnelle.....	60
4. Estimation par la méthode du maximum de vraisemblance d'un modèle ARMA	

périodique	60
4.1. Le modèle et sa fonction de vraisemblance	60
4.1.1. La fonction de vraisemblance quand $p = 0$	61
4.1.2. Approximation de la fonction de vraisemblance pour $p > 0$	66
4.2. Estimation du maximum de vraisemblance	66
4.3. Algorithme	66
5. Applications numériques	67

Deuxième Partie Estimation en-ligne

Chapitre 4 Estimation en-ligne des modèles ARMA classiques

1. Introduction	78
2. Méthode d'erreur de prévision minimale (RPEM)	79
2.1. Principes	79
2.2. Critères	80
2.3. Calcul récursif de l'erreur de prévision	81
3. L'algorithme des moindres carrés récursif (RLS)	82
4. Description de l'algorithme RML (<i>recursive Maximum Likelihood</i>).....	85
4.1. Préliminaires.....	85
4.2. Optimisation stochastique	85
4.3. Approximation de $\hat{\varepsilon}_t(\hat{\beta}_{t-1})$ et $\psi_t(\hat{\beta}_{t-1})$	88
4.4. Estimation récurrente de σ^2	90
5. Algorithme de Mélard-Zahaf (RML _{MZ})	90
5.1. Méthode proposée	90
5.2. Calcul de la matrice d'information de Fisher pour un modèle ARMA(p,q)	91
6. Analyse des propriétés statistiques des algorithmes RML et RML _{MZ}	94
6.1. Généralités	94
6.2. Propriétés statistiques de l'algorithme RML	95
6.3. Propriétés statistiques de l'algorithme RML _{MZ}	95
6.3.1. Préliminaires.....	95
6.3.2. Convergence de l'estimateur RML _{MZ}	95

Chapitre 5 Estimation en-ligne des modèles ARMA périodiques (PARMA)

1. Introduction	97
2. Méthode RPEM dans les modèles ARMA périodiques	98
2.1. Critères	98
2.2. Calcul récursif de l'erreur de prévision	99
3. L'algorithme des moindres carrés récursif (RLS)	100
3.1. Estimation récurrente de σ_i^2 $i = 1, \dots, d$	102
4. Extension de l'algorithme RML au cas périodique	103
4.1. Approximation de $\hat{\varepsilon}_t(\hat{\beta}_{t-1})$ et de $\psi_t(\hat{\beta}_{t-1})$	106
5. Description de l'algorithme RML _{MZ} périodique	107
5.1. Calcul de la matrice d'information de Fisher pour un modèle ARMA périodique.....	107
5.1.1. Calcul de la matrice d'information de Fisher d'un modèle PAR _d (p).....	108
5.1.2. Calcul de la matrice d'information de Fisher d'un modèle PARMA _d (p,q)	109
5.2. Algorithme RML _{MZ}	115
5.3. Propriétés asymptotiques des estimateurs RML et RML _{MZ} périodiques.....	116
5.4. Calcul des autocovariances d'un modèle autorégressif périodique.....	116
5.5. Exemples.....	117

Chapitre 6 Implémentation

1. Introduction	120
2. Choix des conditions expérimentales	121
2.1. Choix des valeurs initiales	121
2.2. Stabilité et admissibilité de la valeur courante $\hat{\beta}_t$	121
2.3. Choix de suite de gains scalaires et facteur d'oubli λ_t	122
2.4. Choix de la suite de gains matriciels R_t	122
3. Simulation et validation	123
3.1. Choix des modèles	123

3.2. Description du programme de simulation.....	123
3.3. Détail des simulations	124
3.3.1. Génération de séries artificielles	124
3.3.2. Estimation récurrente des paramètres	124
3.3.3. Calcul des statistiques de comparaison.....	124
4. Résultats	125
4.1. Résultats concernant les modèles ARMA classiques	125
4.2. Résultats concernant les modèles ARMA périodiques	126
Conclusion	156
Références	158

Introduction générale

1.Introduction

L'analyse des séries chronologiques occupe une place prépondérante dans l'étude des phénomènes aléatoires, à savoir l'économie, l'automatique, la météorologie, l'hydrologie... Celle-ci rassemble un ensemble de techniques et de méthodes mathématiques permettant la construction de modèles stochastiques à travers des séries de données indexée dans le temps. Cette construction se fait généralement selon trois étapes essentielles : l'identification du modèle paramétrique, l'estimation de ses paramètres et les tests permettant la validation de cette construction. Le modèle ainsi obtenu peut être exploité entre autres dans la prévision ou le contrôle.

Bien qu'ils ne représentent souvent qu'une approximation brutale de la réalité, les modèles *stationnaires* (ou ceux qui peuvent être rendus stationnaires) tels que les modèles ARIMA, SARIMA, VARMA, etc., sont les plus répandus et les plus utilisés dans l'analyse des séries temporelles. Depuis les années 70 et avec la parution du fameux livre de Box et Jenkins (1970) ces modèles ont fait l'objet d'une littérature surabondante en effet, les travaux de ces deux auteurs ont rendu célèbres ce genre de modèles, en proposant une méthodologie permettant la construction du modèle adéquat.

Une fois le modèle identifié, *l'estimation* de ses paramètres est sûrement une étape très importante dans la construction de celui-ci. Elle a constitué, jusqu'à lors le centre d'intérêt des chercheurs, et s'est soldée par de nombreux résultats intéressants. Les méthodes d'estimations sont généralement classées en deux grandes classes, et ce, en fonction du type de données sur lesquelles elles se basent : les méthodes *hors-ligne* et les méthodes *en-ligne*. Dans le premier cas, il s'agit d'une série de taille fixe, par contre dans le second les données sont progressivement disponibles.

Les techniques d'estimation *hors-ligne* des paramètres de modèles de séries chronologiques, peuvent être classées en deux catégories générales : les méthodes d'estimation utilisant directement les données et les méthodes qui n'utilisent les données qu'à travers une transformation de celles-ci en résumés statistiques. Parmi les méthodes de la première classe nous citons en particulier la méthode du maximum de vraisemblance et ses différentes approximations (ces approximations sont astreintes à préserver quelques propriétés asymptotiques telles que la convergence et l'efficacité asymptotique) et la méthode des moindres carrés et ses variantes. Parmi les méthodes de la seconde classe, probablement l'approche commune est de transformer les données en un ensemble fini de résumés statistiques (moyenne empirique, covariance empirique) et d'estimer, par la suite, les paramètres du modèle (estimateur de Yule-Walker, estimateur de Durbin-Levinson,...). Pour les paramètres d'un modèle autorégressif moyenne mobile (ARMA), la méthode des *moindres carrés* (conditionnelle et non conditionnelle) a été l'une des premières à être utilisée et a montré des résultats satisfaisants, mais le besoin croissant de l'analyse des séries a eu pour conséquence, la naissance de nouveaux problèmes pour lesquels cette méthode est devenue lourde et insuffisante. Avec l'amélioration des performances des ordinateurs ainsi que l'apparition de nouveaux algorithmes, la méthode du *pseudo maximum de vraisemblance* s'est révélée d'une importance capitale, aussi bien du point de vue théorique que pratique, et a donné beaucoup de satisfaction pas pour la précision des résultats seulement, mais aussi pour la stabilité et la rapidité des calculs. Donc il est tout à fait naturel que cette méthode figure maintenant dans les meilleurs logiciels de statistique. Le principe de celle-ci consiste à optimiser, donc à évaluer plusieurs fois, une fonction non linéaire en inversant à chaque itération une certaine matrice. Plusieurs algorithmes ont vu le jour, afin de palier aux carences dû à l'inversion des matrices, pour justement réduire le temps de calcul et l'espace mémoire.

Le traitement en-ligne des données était au début propre aux problèmes qu'encouraient les ingénieurs, car ils ont été les premiers à les confronter, (les phénomènes à temps continu : traitement de signal en automatique, transmission, télécommunication, etc.). Cependant, avec l'avènement de grandes bases de données plusieurs domaines du fait de leur complexité croissante, font actuellement appel à des données en-ligne (météorologie, économétrie, finances et bourses,...). De ce fait, tout laisse à croire que les méthodes en-ligne, bien qu'elles soient peu répandues en statistique, connaîtrons sans doute un nouvel essor. Les méthodes d'estimations hors-ligne sont trop lourdes voir impossibles lorsqu'elles sont appliquées à des données en-ligne, car leur utilisation contraint, à chaque introduction d'une nouvelle donnée, de refaire la procédure d'estimation pour tout le bloc des données ce qui est très coûteux en terme de complexités spatiale et temporelle. Face à ce problème, les ingénieurs ont développé une méthodologie complète, portant le nom d'*identification récurrente*. Cette méthodologie part de la conception des algorithmes à l'examen des conditions de mise en œuvre numérique sur calculateur, en passant par l'étude approfondie de la convergence et l'optimisation de leur vitesse. Une des raisons de l'existence de plusieurs algorithmes récurrents est la diversité de la nature des problèmes à traiter. Cette diversité implique l'utilisation de différentes approches qui peuvent être énumérées selon les quatre classes suivantes : méthodes issues par modification des méthodes hors-ligne, méthodes basées sur les filtres non linéaires, les méthodes d'approximations stochastiques et les méthodes pseudo-régression linéaire. Les algorithmes récurrents ont occupé une place considérable, dû au développement du traitement numérique et à l'augmentation constante de la puissance des calculateurs, permettant ainsi l'implémentation en temps réel d'algorithmes de plus en plus sophistiqués. L'intérêt de la forme récursive des algorithmes récurrents, est dû à la possibilité du traitement en-ligne des données en temps réel, grâce à une mémorisation finie dans un vecteur de taille fixe. En effet l'estimation courante est remise à jour de manière

à ce que le nombre d'opérations et l'espace mémoire ne croissent pas avec le nombre d'observations, et ne dépendent que du modèle choisi.

Parmi de nombreuses variantes de l'estimation en-ligne, nous avons choisi la classe des méthodes issue de l'identification hors-ligne, appelée méthodes *d'erreur de prévision minimale* ('Recursive Predictor Error Method' RPEM). Dans le cas gaussien, ces méthodes conduisent à la même fonction critère que celle du *maximum de vraisemblance conditionnelle*, et donne des estimateurs de qualités proches. On notera par ailleurs deux cas particuliers de cette méthode : le premier, relatif aux modèles autorégressifs, est l'algorithme des moindres carrés récursifs (RLS), quant au second qui n'est autre que la généralisation du premier, aux cas de l'ajout de composante moyenne mobile, est l'algorithme *Recursive Maximum Likelihood* (RML). Celui-ci donne, sous certaines conditions, des estimateurs ayant des propriétés statistiques (convergence, efficacité asymptotique,...) voisines de celles de l'estimateur du maximum de vraisemblance (Ljung et Soderstrom (1983) et Duflo (1990)). L'idée de base de ces algorithmes est de condenser l'information apportée par l'échantillon dans une quantité auxiliaire, exprimée sous forme d'une matrice, et de remettre à jour l'estimateur, à l'introduction d'une nouvelle donnée. Cependant, Mélard (1989) a observé que cette quantité auxiliaire est fort variable, générant ainsi des perturbations nuisibles à l'estimation. En conservant le principe de l'algorithme RML, Zahaf (1998) y a proposé une amélioration en remplaçant la quantité d'information auxiliaire par son espérance mathématique, qui n'est autre que la matrice d'information de Fisher. Il a, en outre, analysé les propriétés statistiques de l'estimateur obtenu par l'algorithme proposé (RML_{MZ}) et a montré aussi que, sous quelques conditions de régularités moins restrictives, celui-ci converge. Il a ainsi vérifié, suite à une étude de comparaison par simulation, que l'algorithme RML_{MZ} est dans la plupart des cas meilleur que l'algorithme RML.

Dans la pratique, de nombreuses séries chronologiques présentent un caractère non régulier qui ne peut être remédié par des transformations tendanciennes, saisonnières ou mixtes. Utiliser les modèles classiques de Box et Jenkins pour traiter ce genre de séries serait donc insuffisant et inefficace. Il est clair que la manipulation des modèles non stationnaires, bien qu'elle soit plus astreignante et délicate, serait plus intéressante car il existe des phénomènes aléatoires ne pouvant être représentés de manière adéquate par les modèles ARIMA classiques de Box et Jenkins (1976), mais plutôt par des modèles linéaires à *coefficients évolutifs* dans le temps. En effet, la généralisation due à Cramér (1961), de la *décomposition* de Wold (1938), aux processus non stationnaires, a permis d'élargir la classe des modèles linéaires autorégressifs moyenne mobile, à coefficients constants (ARMA), à la classe des modèles linéaires autorégressifs moyenne mobile à coefficients dépendant du temps (ARMA_t).

Une classe particulière des modèles ARMA_t évolutifs qui est souvent très utile pour la modélisation des séries saisonnières est celle dont les paramètres sont périodiques dans le temps. L'incapacité des modèles SARIMA, proposés par Box et Jenkins (1970), à représenter certaines séries saisonnières, spécialement pour les séries issues de processus dont la fonction d'autocovariance est périodique (Tiao et Crupe (1980)), a favorisé les développements apportés aux travaux de recherches relatifs à cette classe de modèles. En effet, plusieurs chercheurs se sont penchés aussi bien sur l'étude des processus *périodiquement corrélés* que celle des modèles *ARMA périodiques*. Pour étudier les propriétés théoriques des modèles ARMA périodiques nous pouvons distinguer deux approches différentes : la première, appelée "perio-span-lumping", se base directement sur le théorème de Gladyshev (1961) et consiste à ramener un processus périodiquement corrélé univarié au

processus stationnaire multivarié qui lui est associé et d'étudier, au sens du second ordre les propriétés de ce dernier. Soulignons que cette technique a été appliquée par plusieurs auteurs afin d'étudier les différents problèmes liés aux modèles linéaires à coefficients périodiques (Pagano (1978), Tiao et Crupe (1980)...). Par la suite d'autres chercheurs ont exploité cette approche, nous citerons entre autres les travaux de Newton (1982), Cipra (1984) et Ghysels et Hall (1992), pour étudier le problème d'inversibilité des modèles univariés périodiques. Cependant, la condition d'inversibilité obtenue par ces auteurs n'est que suffisante ce qui a été montré dans Bentarzi et Hallin (1994). Bentarzi et Hallin (1994) ont élaboré par contre une condition nécessaire et suffisante pour l'inversibilité des modèles moyenne mobile multivariés d -périodiques, d'ordre q . La seconde approche, quant à elle, consiste à ramener l'étude des propriétés d'un modèle m -varié, d -périodique d'ordre q , à l'étude de ces propriétés pour un modèle mq -varié d -périodique d'ordre un. Cette dernière technique, introduite par Bentarzi et Hallin (1994) et appelée par Ula et Smadi (1997) "order-span-lumping", permet l'étude de plusieurs propriétés théoriques d'un modèle m -varié moyenne mobile d -périodique tels que l'inversibilité et la *factorisation spectrale* (voir Bentarzi (1998), Bentarzi et Hallin (1998), Bentarzi (1995) et Bentarzi et Hallin (1994)). En adoptant l'approche "order-span-lumping", pour les processus moyenne mobile univarié d'ordre un (respectivement m -varié d'ordre fini quelconque), d -périodique, le célèbre problème de la factorisation spectrale d'un processus périodique qui consiste à la construction de tous les modèles adéquats possibles correspondants à une fonction d'autocovariances périodique donnée, d'un certain processus périodiquement corrélé, à été résolu dans Bentarzi et Hallin (1998) (respectivement Bentarzi (1998)). Ula et Smadi (1997) ont exploité par la suite cette technique pour introduire par analogie une condition nécessaire et suffisante de causalité des modèles ARMA d -périodiques.

Depuis la fin des années 70, la modélisation ARMA périodique (identification, estimation et validation) n'a pas cessé d'être en expansion. L'objectif principal était de généraliser les méthodes classiques relatives aux modèles ARMA, au cas de modèles ARMA périodiques. En effet le problème de l'identification a été étudiée par plusieurs chercheurs, citons en particulier : Cleveland et Tiao (1979), Sakai (1982), Vecchia (1985b), Mcleod, (1992), etc. La plupart de ces travaux visaient à généraliser les méthodes basées sur la théorie de l'information, tels que les critères : AIC, BIC, CIC ou celles basées sur les fonctions d'autocorrélations, les fonctions d'autocorrélations partielles et les fonctions d'autocorrélations inverses (Benouamer et Bentarzi (1996) pour la méthode du coin). Plus récemment, Hemis (1999) et Bentarzi (2000) ont généralisé une méthode basée sur l'analyse bayésien (critère PDC) introduite par Djuric et Kay, (1992) dans le cas ARMA classique. Dans le même travail, Hemis (1999) s'est intéressée d'une part à l'estimation, par la technique de Gibbs, des paramètres de modèles autoregressifs périodiques à *tendance explicative* et d'autre part à l'étude de l'effet de l'application des estimateurs obtenus, en introduisant la technique de Gibbs, sur la performance des critères de sélection de l'ordre du processus autoregressif pur et à tendance explicative. Le problème de la validation qui consiste essentiellement à tester le modèle obtenu par rapport aux observations n'a pas eu par contre autant d'intérêt, citons exceptionnellement quelques travaux, tels que Vecchia et Balerini, (1991), pour le test de la périodicité et Mcleod, (1992) pour le problème de la validation mais en se restreignant uniquement dans les modèles AR périodiques. En étudiant la propriété de normalité asymptotique locale (LAN) ainsi que la propriété de linéarité asymptotique (Bentarzi (1995), Bentarzi et Hallin (1996)), Bentarzi (1995) a construit un test (paramétrique et non paramétrique) localement asymptotiquement le plus serré ("stringent") pour tester un modèle autoregressif classique (hypothèse nulle) contre un modèle autoregressif périodique

(hypothèse alternative locale contiguë) (voir aussi Bentarzi et Hallin (1993) et Bentarzi et Hallin (1996)).

Pour des données hors-ligne, les méthodes d'estimation des modèles périodiques ont connu un développement important. La littérature est en fait surabondante et l'intérêt n'a pas cessé de croître. Après la généralisation (Cramér (1961)), aux processus non stationnaires, du théorème de Wold (1938) l'étude des processus périodiquement corrélés (périodiquement stationnaires) amorcée par Gladyshev (1961) a permis à plusieurs chercheurs de s'intéresser aux modèles périodiques. John et Brelford (1967), en traitant des modèles à structure périodique, se sont penchés sur le problème de l'estimation des paramètres de modèles autorégressifs périodiques particuliers mais sans pour autant étudier les propriétés asymptotiques. Après un nombre considérable de travaux n'ayant pas abouti à des résultats consistants, Pagano (1978) a obtenu des résultats fort-intéressants concernant les modèles AR périodiques. Il a en outre, déterminé les estimateurs des moments, établi leurs propriétés statistiques et a généralisé les estimateurs de Yule-Walker tout en assurant leur convergence. D'autres parts des chercheurs, tel que Troutman (1979) et Cleveland et Tiao (1979) ont traité le même cas par l'usage d'approches différentes de celles déjà utilisées. Grâce à Tiao et Crupe (1980), qui ont montré l'insuffisance des modèles saisonniers de Box et Jenkins (1976) à représenter les modèles à caractère périodique, la modélisation ARMA périodiques, qui était d'usage courant dans le domaine de l'hydrologie, a connu un développement important, notamment, les problèmes de l'estimation des paramètres. Depuis, une large série de travaux concernant l'estimation a été effectués par plusieurs chercheurs, citons en particulier, Salas et *al.* (1982), Vecchia et *al.* (1983), Vecchia (1983) et Andel (1983). Par ailleurs, Sakai (1982) a amélioré du point de vue algorithmique la méthode de Pagano, (1978) en proposant une approche récursive (vis-à-vis de l'ordre du modèle) basée sur la généralisation, aux modèles périodiques, des équations récursives de Durbin-Levinson, il a, de plus généralisé la notion d'autocorrélation partielle, qui se trouve être un instrument très utile dans les problèmes de l'identification. Jusque là, la plus part des travaux et investigations de l'estimation des paramètres se sont principalement axés sur les modèles AR périodiques. Cipra, (1985), par contre, s'est intéresser aux modèles moyenne mobile périodiques, en adaptant les méthodes de Durbin (1959) consistant à transformer un modèle moyenne mobile à un modèle AR d'ordre élevé aux modèles périodiques. Indépendamment des travaux antérieurs, Vecchia (1985) a proposé un algorithme efficace basé sur le critère du maximum de vraisemblance, la méthode proposée donne des estimateurs exacts pour le cas de modèles MA et approximatifs pour le cas de l'ajout d'une composante AR. Pour la même approche, Li et Hui (1988) ont proposé un algorithme de calcul de la fonction de vraisemblance exacte d'un modèle ARMA périodique. Jiminèz et *al.* (1989) ont par la suite introduit un algorithme basé sur le filtre de Kalman. Plus récemment, Boshnakov (1994a, 1994b) s'est intéressé aux méthodes récursives (vis-à-vis de l'ordre) d'estimation de modèles AR périodiques, et a aussi généralisé (Boshnakov, (1996)) la méthode de Sakai (1982) et la méthode de Franke (1985) aux modèles PARMA. L'estimation en-ligne n'a pas connue par contre autant de succès dans le cas de modèles ARMA périodique, la littérature est en fait relativement maigre par rapport à l'estimation hors-ligne. Adams et Goodwin (1995) ont introduit un algorithme récursif pour l'estimation en-ligne d'un modèle PARMA en généralisant un algorithme de la classe des méthodes pseudo régression linéaire (PLR), ils ont établi la convergence et ont montré aussi l'efficacité numérique de l'algorithme proposé. Dans le cas adaptatif, Talbi et Bentarzi (1996) ont exploité la propriété LAN et la propriété de linéarité et la propriété Minimax, au sens de Fabian et Hannan (1982) pour construire des estimateurs Localement Asymptotiquement Minimax Adaptatifs (LAM-Adaptif) pour l'estimation des paramètres des modèles autorégressifs périodiques.

L'intérêt du travail proposé sera porté sur l'estimation, en présence des données en ligne et hors ligne, d'un modèle autorégressif moyenne mobile périodique (PARMA). Nous tenterons, donc d'étendre et de généraliser quelques méthodes d'estimation en-ligne et hors ligne aux modèles PARMA. Dans un premier temps, nous faisons une extension, aux modèles PARMA périodiques, de l'algorithme RML. L'application de l'algorithme RML_{MZ} aux modèles ARMA stationnaires, ayant donné, des résultats meilleurs que ceux obtenus par l'algorithme RML, nous laisse à croire que la supériorité de RML_{MZ} reste valable pour le cas des modèles PARMA. Dans cette optique, nous essayerons de généraliser l'algorithme RML_{MZ} pour la classe des modèles PARMA. Nous aurons donc besoin, de calculer la matrice d'information de Fisher pour un modèle PARMA, un algorithme spécifique basé sur l'algorithme de Klein et Mélard (1989) sera introduit. Notons que ce dernier fait appel au calcul des autocovariances d'un certain modèle autorégressif périodique, pour cela nous avons élaboré une procédure basée sur l'algorithme de McLeod (1975). Nous vérifierons finalement, par une étude de simulation intensive, les résultats obtenus.

2. Présentation et apport de la thèse

Ce travail intitulé estimation en-ligne et hors-ligne des modèles ARMA périodiques, est bâti sur deux parties essentielles. La première, se composant de trois chapitres, concerne l'étude des problèmes de l'estimation hors-ligne, dans les modèles ARMA classiques et dans les modèles ARMA périodiques. La seconde partie (en trois chapitres), quant à elle, est consacrée aux méthodes d'estimation en-ligne dans les deux classes de modèles : ARMA classique et ARMA périodiques. Dans ce qui suit, nous exposerons brièvement les problèmes majeurs étudiés et les contributions essentielles apportées.

Première Partie Estimation hors-ligne

La première partie de notre travail sera consacrée à l'estimation en présence de données fixes de modèles ARMA classiques et périodiques.

Chapitre 1

Puisque nous traitons de l'estimation en-ligne et hors-ligne dans les modèles PARMA, il sera indispensable de rappeler dans un premier chapitre les principaux résultats concernant l'estimation dans les modèles ARMA classique. Nous rappelons les critères essentiels de base pour l'estimation ainsi que les méthodes d'optimisations utilisés, nous passerons en revue les algorithmes les plus connus.

Chapitre 2

Le deuxième chapitre du travail sera consacré à l'étude des processus périodiquement corrélés et les modèles autorégressifs moyenne mobile périodique, nous y présenterons les méthodes et résultats théoriques principaux et y passerons en revue les propriétés de base concernant ces méthodes.

Chapitre 3

Le troisième chapitre sera consacré à l'estimation hors-ligne dans les modèles ARMA périodique (PARMA). Après avoir rappeler les critères utilisés, nous exposerons deux classes de méthodes d'estimations : la première concerne les méthodes qui utilisent des transformations de données en résumés statistiques (fonction d'autocovariance), basée sur les équations de Yule-Walker périodiques, ce genre de méthodes est souvent appelé méthode des moments. La seconde classe concerne les méthodes (itératives en général) qui utilisent directement les données, nous distinguons en particulier, la méthode des moindres carrés et

ses variantes et la méthode du maximum de vraisemblance et ses approximations. Compte tenu du caractère itératif de ce type de méthodes, de bonnes valeurs initiales seront nécessaires, les estimations par la méthode des moments en seront un bon moyen. Nous terminons le chapitre par une application numérique permettant de comparer entre certaines méthodes en faisant ressortir la meilleur.

2^{ème} Partie Estimation en-ligne

Cette deuxième partie traite de l'estimation en présence de données en-ligne de modèles ARMA classiques et périodiques.

Chapitre 4

Dans le quatrième chapitre, nous nous intéressons à l'estimation hors-ligne dans les modèles ARMA classiques. Nous présenterons donc la méthode de l'erreur de prévision récurrente (*Recursive Predictor Error Method*, RPEM) ainsi que son cas particulier RML sur un modèle autorégressif moyenne mobile de paramètres p et q ARMA(p,q). Nous détaillerons en suite l'algorithme RML_{MZ} (amélioration de Mélard et Zahaf) basé sur le calcul de la matrice d'information de Fisher pour un modèle ARMA(p,q). Nous terminerons le chapitre par l'étude des propriétés théoriques de ces algorithmes.

Chapitre 5

Nous nous intéresserons dans le cinquième chapitre à l'estimation en-ligne d'un modèle autorégressif moyenne mobile périodique PARMA_d(p,q). Nous présenterons en premier lieu une extension de l'algorithme des moindres carrés récurrents (RLS) à la classe des modèles PARMA, nous proposons ensuite une généralisation de l'algorithme RML pour cette même classe de modèles. Nous passerons ensuite à la formulation de l'algorithme RML_{MZ} pour les modèles PARMA. Notons qu'à chaque itération de cet algorithme il est nécessaire d'obtenir la matrice d'information de Fisher des paramètres du modèle PARMA ; un algorithme spécifique sera présenté. Nous terminons le chapitre par l'exposé d'une méthode d'obtention des autocovariances d'un modèle PAR ; celle-ci sera indispensable pour l'obtention de la matrice d'information de Fisher.

Chapitre 6

Ce chapitre est consacré aux problèmes de l'implémentation numérique sur ordinateur des algorithmes récurrents que nous avons établit, nous montrerons sur des séries artificielles, générées par simulation, l'efficacité des algorithmes proposés.

Première Partie

Estimation hors-ligne

Chapitre 1

Estimation hors-ligne dans les modèles ARMA classiques

1. Introduction

Tous les problèmes d'estimation évoqués dans ce travail concernent l'estimation paramétrique, c'est à dire le cas où la famille de lois de probabilité de la composante aléatoire du modèle sous-jacent est connue (généralement supposée gaussienne). Le but de ce chapitre est de donner un aperçu général sur l'estimation hors-ligne des paramètres d'un modèle ARMA.

Le plan de ce chapitre est alors le suivant : Nous aurons besoin de commencer en premier lieu par la présentation des modèles ARMA classiques (autorégressifs moyenne mobile). Nous rappelons entre autres quelques propriétés théoriques essentielles telles que la causalité et l'inversibilité, nous présentons en particulier les équations de Yule-Walker pour les modèles autorégressifs purs. Nous exposerons ensuite la classe des méthodes utilisant des

transformations des données et qui est souvent appelée méthode des moments. Ce type de méthodes est basé sur les équations de Yule-Walker. Nous passons après, à l'algorithme de Durbin-Levinson pour l'estimation récursive des paramètres d'un modèle autorégressif qui, lui aussi, est basé sur ces équations. Nous détaillerons ensuite l'algorithme de Wilson (1979) pour le calcul récursif des fonctions d'autocovariances qui n'est rien d'autre qu'une variante de l'algorithme de Durbin et qui sera utilisé tout au long de ce travail. La quatrième section sera consacrée à l'étude des méthodes utilisant directement les données, comme la plupart de ces méthodes sont itératives nous serons ramener à exposer aussi quelques routines d'optimisation que nous aurons besoin dans la suite de notre travail. Nous terminons ainsi le chapitre par un exposé des propriétés limite des estimateurs étudiés.

2. Généralités

Tout au long de ce travail, nous considérons un processus, $\{y_t, t \in Z\}$, du second ordre, centré.

Définition 1.2.1 (Modèle autorégressif moyenne mobile)

On dit que le processus aléatoire $\{y_t, t \in Z\}$ admet une représentation autorégressive moyenne mobile d'ordre p et q notée ARMA(p, q), s'il est solution de l'équation aux différences stochastique suivante :

$$y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \dots - \phi_p y_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}. \quad (1.2.1)$$

où $\{\varepsilon_t, t \in Z\}$ est un processus bruit blanc de variance σ^2

L'équation (1.2.1) est équivalente à la suivante :

$$\phi(L)y_t = \theta(L)\varepsilon_t$$

où $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$ est un polynôme d'ordre p , $\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q$ est un polynôme d'ordre q et L est l'opérateur retard défini par : $L^j y_t = y_{t-j}$, $j = 0, 1, \dots, p$.

Remarque

- Le modèle particulier obtenu à partir de (1.2.1), lorsque les (θ_j) , $j = 1, \dots, q$ sont nuls, est appelé modèle autorégressif d'ordre p que l'on note par AR(p).

- De même, lorsque les (ϕ_j) , $j = 1, \dots, p$ sont nuls le modèle particulier obtenu à partir de (1.2.1), est appelé modèle moyenne mobile d'ordre q .

3. Estimation par la méthode des moments

Avant de passer à l'étude des méthodes des moments pour l'estimation des paramètres d'un modèle AR, il nous semble utile de rappeler les équations qui sont de base pour ces méthodes et qui permettent d'établir un lien entre les paramètres d'un modèle autorégressif pur et ses autocovariances. Nous présentons dans un premier temps, deux méthodes d'estimation (dites méthodes des moments) qui utilisent des transformations des données. La première, appelée méthode de Yule-Walker, la seconde, quant à elle, est une amélioration de la méthode de Yule-Walker du point de vue algorithmique et est appelé méthode de Durbin-Levinson, notons que ces méthodes concernent uniquement les modèles autorégressifs purs. Pour le cas ARMA général le système de Box et Jenkins, (1976) substituera la méthode de Yule-Walker. Quant à la méthode récursive de Durbin-Levinson elle sera remplacée par la méthode de Franke, (1985).

3.1. Les équations de Yule-Walker

Considérons le processus $\{y_t, t \in Z\}$, autorégressif causal, suivant :

$$y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \dots - \phi_p y_{t-p} = \varepsilon_t. \quad (1.3.1)$$

Sur la base d'une observation y_1, y_2, \dots, y_N , supposée être une réalisation du processus (1.3.1), nous désirons estimer le vecteur des paramètres $\phi^T = (\phi_1, \phi_2, \dots, \phi_p)$ et la variance du bruit blanc σ^2 . En multipliant chaque membre de l'équation (1.3.1) par y_{t-j} , $j = 0, \dots, p$, et en prenant l'espérance des membres obtenus nous aurons un système d'équations appelé les équations de Yule-Walker, d'où

$$\begin{aligned} \Gamma_p \phi &= \gamma_p \\ \sigma^2 &= \gamma_0 - \phi^T \gamma_p, \end{aligned} \quad (1.3.2)$$

où Γ_p est la matrice des covariances $(\gamma_{i-j})_{i,j=1,\dots,p}$ et $\gamma_p = (\gamma_1, \gamma_2, \dots, \gamma_p)^T$

Ces équations, comme nous allons voir, peuvent être également utilisées pour déterminer les autocovariances $\gamma_0, \gamma_1, \dots, \gamma_p$ à partir de la connaissance de σ^2 et de ϕ .

3.2. Estimateurs de Yule-Walker (des moments)

Par ailleurs, si nous remplaçons les autocovariances théoriques γ_j , $j = 0, 1, \dots, p$, par leurs estimations empiriques $\tilde{\gamma}_j$, nous obtenons un ensemble d'équations permettant d'obtenir les estimateurs $\tilde{\phi}$ et $\tilde{\sigma}^2$ appelés estimateurs de Yule-Walker de ϕ et σ^2 .

$$\begin{aligned} \tilde{\Gamma}_p \tilde{\phi} &= \tilde{\gamma}_p, \\ \tilde{\sigma}^2 &= \tilde{\gamma}_0 - \tilde{\phi}^T \tilde{\gamma}_p, \end{aligned} \quad (1.3.3)$$

où $\tilde{\Gamma}_p$ est la matrice des covariances empiriques : $(\tilde{\gamma}(i-j))_{i,j=1,\dots,p}$ et $\tilde{\gamma}_p = (\tilde{\gamma}_1, \tilde{\gamma}_2, \dots, \tilde{\gamma}_p)^T$.

Remarques

- Ces estimateurs sont aussi appelés les estimateurs des moments, puisqu'ils sont obtenus au moyen du remplacement des moments théoriques par leurs estimations empiriques.
- Souvent, les estimateurs des moments sont moins efficaces que les estimateurs obtenus par des méthodes alternatives tels que la méthode des moindres carrés ou la méthode du maximum de vraisemblance. Cependant, pour un modèle AR(p), il est montré (voir Brockwell et Davis (1988)) que les estimateurs de Yule-Walker possèdent des propriétés asymptotiques équivalentes à celle de l'estimateur du maximum de vraisemblance.

Théorème 1.2.1 (Convergence)

Si $\{y_t, t \in Z\}$ est un processus, autorégressif causal, AR(p) et $\tilde{\phi}$ est l'estimateur de Yule-Walker du vecteur des paramètres ϕ , alors

$$N^{1/2}(\tilde{\phi} - \phi) \rightarrow N(0, \sigma^2 \Gamma_p^{-1}), \quad (1.2.4)(1.3.4)$$

où, Γ_p est la matrice des covariances $(\gamma(i-j))_{i,j=1,\dots,p}$.

4. Algorithmes récursifs pour l'estimation des moments

Supposons que nous disposons d'une observation y_1, y_2, \dots, y_N d'une série stationnaire. Nous admettons que cette série est une réalisation d'un processus stochastique satisfaisant un modèle autorégressif d'ordre p défini par (1.2.1). Alors nous envisageons les deux problèmes suivants : Le premier consiste à estimer les paramètres $\phi_1, \phi_2, \dots, \phi_p$ sur la base de la connaissance des autocovariances γ_h via les équations de Yule-Walker. Le second quant à lui (problème inverse) consiste à déterminer les autocovariances γ_h sur la base de la connaissance des paramètres du modèle.

Nous allons donc donner des équations récursives permettant d'établir le lien entre les paramètres du modèle et ses autocovariances. L'idée générale de cette méthode se réduit à construire une suite de modèles AR(k) d'ordre k , pour k variant de 1 jusqu'à p , qui *ajustent* pour chaque rang les données selon un critère déterminé (au sens des moindres carrés).

4.1. Algorithme de Durbin pour l'estimation des paramètres

Soit y_1, y_2, \dots, y_N une série d'observations d'un processus autorégressif d'ordre p . Alors le schéma récursif pour calculer les estimateurs $\tilde{\phi}_1, \tilde{\phi}_2, \dots, \tilde{\phi}_p$ de $\phi_1, \phi_2, \dots, \phi_p$ à partir de $\tilde{\gamma}(0), \tilde{\gamma}(1), \dots, \tilde{\gamma}(p)$ est donné par l'algorithme de Durbin (1960) suivant :

(Pour alléger l'écriture nous allons omettre le symbole tilde qui symbolise l'estimateur des paramètres et non pas les paramètres.)

Algorithme 1.1**Etape (1) (Initialisation)**

Commencer par calculer les valeurs initiales σ_0^2 et ϕ_{11} au moyen des équations :

$$\begin{cases} \sigma_0^2 = \gamma_0 \\ \alpha_1 = \phi_{11} = \gamma_1 / \sigma_0^2 \end{cases} \quad (1.4.2)$$

Etape (2) (Construction des $\phi_{k,j}$)

Pour $k = 0, \dots, p-1$, nous appliquons les équations suivantes dans l'ordre :

$$\alpha_{k+1} = \phi_{k+1,k+1} = (\gamma_{k+1} - \phi_{k,1}\gamma_k - \dots - \phi_{k,k}\gamma_1) / \sigma_k^2 \quad (1.4.3)$$

$$\phi_{k+1,j} = \phi_{k,j} - \alpha_{k+1}\phi_{k,k+1-j} \quad j = 1, \dots, k, \quad (k \geq 1) \quad (1.4.4)$$

$$\sigma_{k+1}^2 = \sigma_k^2 (1 - \alpha_{k+1}^2) \quad (1.4.5)$$

Etape (3)

Poser $\sigma^2 = \sigma_p^2$, $\phi_k = \phi_{p,k}$, $k = 1, \dots, p$

4.3. Algorithme de Wilson pour l'estimation des autocovariances pour un modèle ARMA**4.3.1. Cas autorégressif AR**

Cet algorithme, dû à Wilson (1979), n'est rien d'autre qu'une procédure inverse de l'algorithme de Durbin (1960). Son objectif est de construire, de façon récursive, les autocovariances du modèle à partir de ses paramètres. Supposons alors que les paramètres, $\phi_1, \phi_2, \dots, \phi_p$, σ^2 du modèle sont connus. Nous allons appliquer une procédure inverse à celle de Durbin (1960).

Algorithme 1.2**Etape (1) (Initialisation)**

Poser $\sigma_p^2 = \sigma^2$ et $\phi_{p,k} = \phi_k$, $k = 1, \dots, p$

Etape (2) (Reconstitution des $\phi_{k,j}$)

Pour $k = p-1, \dots, 0$ appliquer les équations suivantes dans l'ordre :

$$\alpha_{k+1} = \phi_{k+1,k+1} \quad (1.4.6)$$

$$\phi_{k,j} = (\phi_{k+1,j} - \alpha_{k+1}\phi_{k+1,k+1-j}) / (1 - \alpha_{k+1}^2), \quad j = 1, \dots, k, \quad (k \geq 1) \quad (1.4.7)$$

$$\sigma_k^2 = \sigma_{k+1}^2 / (1 - \alpha_{k+1}^2). \quad (1.4.8)$$

Etape (3)

Poser $\gamma_0 = \sigma_0^2$

pour $k = 0, \dots, p-1$, faire $\gamma_{k+1} = \phi_{k,1}\gamma_k + \dots + \phi_{k,k}\gamma_1 + \sigma_k^2\alpha_{k+1}$; (1.4.9)

Remarques

1- L'algorithme précédent permet de déterminer les autocovariances uniquement pour les horizons $k = 0, 1, \dots, p$, car ce sont les seuls dont nous aurons besoin ultérieurement (dans l'évaluation de la matrice d'information de Fisher). Bien entendu, pour le cas où $k > p$, nous utiliserions la relation

$$\gamma_{k+1} = \phi_1 \gamma_k + \dots + \phi_p \gamma_{k+1-p}. \quad (1.4.10)$$

2- Lorsque cette procédure inverse est appliquée à un polynôme arbitraire, alors la condition de causalité (stationnarité) est équivalente à $|\alpha_{k+1}| < 1$, $k = 0, \dots, p-1$, voir aussi par exemple Anderson (1975) et Duffin (1969).

3- Le point principal tiré de la construction des $\gamma_0, \gamma_1, \dots, \gamma_p$ à partir de $\phi_1, \phi_2, \dots, \phi_p$, σ^2 par l'utilisation de la procédure inverse est que cette procédure requiert un nombre d'opérations de l'ordre de $O(p^2)$ contrairement à la méthode proposée par McLeod, (1975) qui nécessite la résolution d'un système de p équations et qui requiert par suite un nombre d'opérations élémentaires de l'ordre de $O(p^3)$.

4.3.2. Cas autorégressif moyenne mobile ARMA

Pour un modèle ARMA, le calcul de la fonction d'autocovariances peut être établi de la même manière que celle pour un modèle AR, mais nous allons utiliser une approche alternative basée sur la fonction génératrice.

Considérons le modèle ARMA de (1.2.1) et supposons que $\phi(z) \neq 0$ pour tout $z \in C$ tel que $|z| \leq 1$. Alors la fonction génératrice des autocovariances $\Gamma(z) = \sum_{-\infty}^{+\infty} \gamma_k z^k$ vérifie :

$$\Gamma(z) = \sigma^2 \theta(z)\theta(z^{-1}) / \phi(z)\phi(z^{-1}) \quad (1.4.11)$$

alors le second membre de (1.4.11) se factorise comme suit

$$\theta(z)\theta(z^{-1}) / \phi(z)\phi(z^{-1}) = v(z) / \phi(z) + v(z^{-1}) / \phi(z^{-1}) \quad (1.4.12)$$

où $v(z) = v_0 + v_1 z + \dots + v_r z^r$ et $r = \max(p, q)$. les coefficients de $v(z)$ sont déterminés en résolvant les équations linéaires obtenues par identification des coefficients de z^k , $k = 0, 1, \dots, r$ dans l'identité suivante :

$$\phi(z)v(z^{-1}) + \phi(z^{-1})v(z) = \theta(z)\theta(z^{-1}) \quad (1.4.13)$$

Pour obtenir les γ_k , nous développons $v(z)\phi(z) = g_0 + g_1 z + g_2 z^2 + \dots$ par

$$g_k = \phi_1 g_{k-1} + \dots + \phi_p g_{k-p} + v_k, \quad k = 0, 1, 2, \dots \quad (1.4.14)$$

En prenant $g_k = 0$ pour $k < 0$ et $v_k = 0$ pour $k > r$. Alors nous aurons finalement,

$$\gamma_0 = 2g_0\sigma^2 \quad \text{et} \quad \gamma_k = g_k\sigma^2 \quad \text{pour} \quad k > 0. \quad (1.4.15)$$

Cette approche est économique pour le temps de calcul et pour l'espace mémoire aussi, mais pour déterminer les v_k il faut résoudre un système à r équations. Cependant, Wilson (1979) a proposé une solution récursive qui requiert un nombre d'opérations de l'ordre de $O(r^2)$. Nous ne détaillerons pas donc cette approche, pour une lecture plus complète voir Wilson (1979). Nous allons juste synthétiser cet algorithme.

Algorithme 1.3 Etant donné le modèle ARMA défini par (1.2.1) (nous supposons que $p > q$). Admettons que les paramètres ϕ_i , $i = 1, \dots, p$, θ_j , $j = 1, \dots, q$, et σ^2 sont connus et que nous désirons déterminer les γ_h , $h = 0, 1, 2, \dots$. Alors nous appliquons les cinq étapes suivantes :

Etape (1) Déterminer les coefficients B_k de $B(z) = \sum_{k=-q}^q B_k z^k$ par la relation

$$B(z) = \theta(z)\theta(z^{-1}) \quad \text{ou} \quad \text{également} \quad B_k = B_{-k} = \sum_{i=0}^{q-k} \theta_i \theta_{i+k} \quad \text{pour} \quad k = 0, 1, \dots, q,$$

soit $B_k = 0$ pour $k = q+1, \dots, p$,

Etape (2) (Calcul des coefficients ϕ_{ij} , $i, j = 1, \dots, p$.)

$$\text{Poser} \quad \phi_{p,k} = \phi_k, \quad k = 1, \dots, p$$

$$\text{Poser} \quad \alpha_{k+1} = \phi_{k+1,k+1} \quad \text{pour} \quad k = p-1, \dots, 0$$

$$\phi_{k,j} = (\phi_{k+1,j} - \alpha_{k+1} \phi_{k+1,k+1-j}) / (1 - \alpha_{k+1}^2), \quad (k \geq 1) \quad \text{Pour} \quad j = 1, \dots, k,$$

Etape (3) (Calcul des coefficients B_{ij} , $i, j = 1, \dots, p$.)

$$\text{Pour} \quad k = 1, \dots, p \quad \text{faire} \quad B_{p,j} = B_j$$

$$\delta_{k+1} = B_{k+1,k+1} \quad (1.4.16)$$

$$\text{Pour} \quad j = 1, \dots, k, \quad \text{faire} \quad B_{k,j} = B_{k+1,j} + \delta_{k+1} \phi_{k,k+1-j} \quad \text{avec} \quad k \geq 1 \quad (1.4.17)$$

Etape (4) (Calcul des coefficients v_j , $j = 1, \dots, p$.)

$$\text{Poser} \quad v_{0,0} = \frac{1}{2} B_{0,0}, \quad v_{k,k+1} = \delta_{k+1} \quad k = 0, 1, \dots, p-1, \quad (1.4.19)$$

$$v_{k+1,j} = (v_{k,j} + \alpha_{k+1} v_{k,k+1-j}) / (1 - \alpha_{k+1}^2), \quad j = 0, 1, \dots, k+1 \quad (1.4.20)$$

$$\text{Poser} \quad v_j = v_{p,j}, \quad j = 1, \dots, p,$$

Etape (5) (Calcul des coefficients g_k , $k = 0, 1, 2, \dots$)

$$g_k = \phi_1 g_{k-1} + \dots + \phi_p g_{k-p} + v_k \quad (1.4.21)$$

$$g_k = 0 \quad \text{si} \quad k < 0 \quad \text{et} \quad v_k = 0 \quad \text{pour} \quad k > r.$$

Etape (6) (Calcul des fonctions γ_h , $h = 0, 1, 2, \dots$)

$$\begin{aligned} \gamma_0 &= 2g_0\sigma^2 \\ \gamma_k &= g_k\sigma^2 \quad k > 0 \\ \gamma_{-k} &= \gamma_k. \end{aligned} \tag{1.4.22}$$

5. Méthodes alternatives d'estimation dans les modèles ARMA

Nous passerons ensuite à une description générale des méthodes de la seconde classe qui nécessite un traitement direct des données, en outre la méthode des moindres carrés conditionnelle et non conditionnelle, et la méthode du maximum de vraisemblance non conditionnelle. Notons que dans le cas AR pur les deux classes deux méthodes possèdent des propriétés statistiques équivalentes. Dans le cas ARMA général, il est montré (voir Porat et Friedlander, (1986a)) que les méthodes à traitement direct sont plus efficaces. Comme la plupart des algorithmes d'estimation utilisés dans ce travail sont de nature itérative, nous aurons besoin d'énumérer les différents types de routines d'optimisation que nous utiliserons ultérieurement. Nous terminons le chapitre par l'exposé des propriétés statistiques désirables des différents estimateurs utilisés.

5.1. Critères

Le but est d'estimer les paramètres d'un processus ARMA(p, q) à partir d'une réalisation de taille N . nous commençons tout d'abord par la sélection d'un critère de choix qui caractérise une certaine perte engendrée par l'utilisation d'un tel ou tel estimateur. Nous avons envisagé les trois critères suivants (Box and Jenkins (1976)) :

- Le critère des moindres carrés conditionnel (MCC).
- Le critère des moindres carrés non conditionnel (MCN).
- Le critère du maximum de vraisemblance (MV).

Nous illustrerons essentiellement sur le processus ARMA(p, q) de (1.2.1)

$$y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \dots - \phi_p y_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}. \tag{1.5.1}$$

Supposons que nous voulons estimer les paramètres de ce modèle en se basant sur la réalisation donnée. Alors ces estimateurs seront choisis de façon à minimiser la fonction critère fixée. Comme généralement cette fonction est non linéaire donc nous devons utiliser des routines d'optimisation itératives pour la minimiser, par conséquent plusieurs évaluations seront nécessaires. Il faut d'abord se pencher sur l'évaluation de la fonction critère. Le problème d'optimisation sera abordé dans le dernier paragraphe de cette section.

5.2. Evaluation des fonctions critères

Pour une réalisation y_1, y_2, \dots, y_N le modèle (1.5.1) peut se mettre sous la forme suivante :

$$Ay_N = B\varepsilon_N + CI_* \tag{1.5.2}$$

où $y_N = (y_1, y_2, \dots, y_N)^T$, $\varepsilon_N = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N)^T$ et $I_* = (y_0, \dots, y_{1-p}, \varepsilon_0, \dots, \varepsilon_{1-q})^T$.

Notons que A et B sont des matrices triangulaires inférieures de taille $(N \times N)$ à diagonale unité. La matrice C , quant à elle, est de taille $(N \times (p + q))$ où seulement les $h (= \max\{p, q\})$ premières lignes sont non nulles. L'expression exacte de ces matrices est donnée par :

$$\begin{aligned}
 A_{(N \times N)} &= \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ -\phi_1 & 1 & 0 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ -\phi_p & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & -\phi_p & \cdots & -\phi_1 & 1 \end{pmatrix}, & B_{(N \times N)} &= \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ -\theta_1 & 1 & 0 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ -\theta_p & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & -\theta_p & \cdots & -\theta_1 & 1 \end{pmatrix}, \\
 C_{(N \times (p+q))} &= \begin{pmatrix} -\phi_1 & \cdots & \cdots & -\phi_p & \theta_1 & \cdots & \cdots & \theta_q \\ 0 & -\phi_1 & \cdots & -\phi_{p-1} & 0 & \theta_1 & \cdots & \theta_{q-1} \\ \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\phi_1 & \vdots & \ddots & \ddots & \theta_1 \\ \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 0 & \cdots & \cdots & 0 \end{pmatrix} \tag{1.5.3}
 \end{aligned}$$

5.2.1. Evaluation de la somme des carrés conditionnelle

La méthode des moindres carrés conditionnelle (MCC) basée sur la somme des carrés conditionnelle consiste à calculer

$$S_C(\phi, \theta) = \sum_{t=1}^N \varepsilon_t^2(\phi, \theta)$$

où $\varepsilon_t(\phi, \theta)$ n'est autre que ε_t calculée par récurrence, avec une valeur initiale I_* fixée égale à 0. Nous pouvons donc déterminer à partir de (1.5.1) $\varepsilon_1 = y_1$. puis

$$\varepsilon_t = y_t - \phi_1 y_{t-1} - \cdots - \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \tag{1.5.4}$$

pour $t \geq h$ ($h = \max(p, q)$)

avec p et q fixés, nous pouvons ainsi, expliciter la valeur de ε_t seulement en fonction de ϕ , θ et de y et ce, en substituant les valeurs de chaque membre du membre gauche de l'équation (1.5.4) par sa valeur récursive. Par suite nous obtenons

$$\varepsilon_t(\phi, \theta) = f_t(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, y_1, \dots, y_{t-1}) \tag{1.5.5}$$

où $f_t(\cdot)$ est une fonction dépendante du modèle dont la forme est bien connue dès que nous fixons p et q . La somme des carrés (S_C) conditionnelle à $I_* = 0$ s'écrit donc

$$S_C(\phi, \theta) = \sum_{t=1}^N \varepsilon_t^2(\phi, \theta) = \sum_{t=1}^N f_t(\phi, \theta, y) \quad (1.5.6)$$

Nous remarquons que la fonction $f_t(\cdot)$ n'est pas une fonction quadratique en ϕ et θ . L'équation normale permettant le calcul de ϕ et θ n'est donc pas linéaire.

5.2.2. Evaluation de la somme des carrés non conditionnelle

La méthode de Box et Jenkins, (1976) consiste à déterminer une valeur de I_* meilleure que 0. Si le modèle (1.5.1) était correct, nous pouvons aussi écrire

$$y_t - \phi_1 y_{t+1} - \dots - \phi_p y_{t+p} = \varepsilon'_t - \theta_1 \varepsilon'_{t+1} - \dots - \theta_q \varepsilon'_{t+q} \quad (1.5.7)$$

où les $\{\varepsilon'_t\}$ sont les innovations en temps inversé (Box and Jenkins (1976)). En supposant $\varepsilon'_{N+k} = 0$ pour $k \geq 1$, nous déterminons successivement ε'_N puis $\varepsilon'_{N-1} \dots$

De façon similaire à (1.5.5) ce qui fournit finalement une estimation, \hat{I}_* , pour I_* (prévision rétrospective). En considérant maintenant \hat{I}_* nous pouvons obtenir les $\varepsilon_t(\phi, \theta)$ par (1.5.4) seulement en fonction de ϕ , θ , \hat{I}_* et de y . Nous en déduisons alors la somme des carrés non conditionnelle.

$$S_N(\phi, \theta) = \sum_{t=1}^N \varepsilon_t^2(\phi, \theta) \quad (1.5.8)$$

5.2.3. Evaluation de la fonction de vraisemblance conditionnelle

Nous avons besoin d'une hypothèse très précise sur la distribution liée des $\{\varepsilon_t\}$. Ce sont des variables aléatoires indépendantes, de loi $N(0, \sigma^2)$. Considérons d'abord la fonction de vraisemblance $l_C(\phi, \theta)$ conditionnelle à $I_* = 0$. Alors la fonction densité du vecteur $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N)^T$ est donnée par la relation

$$f(y) = (2\pi)^{-N/2} \sigma^{-N} \exp\left\{-\frac{1}{2\sigma^2} \sum_{t=1}^N \varepsilon_t^2\right\}. \quad (1.5.9)$$

Notons par $y = (y_1, y_2, \dots, y_N)^T$ le vecteur des observations. La matrice Jacobienne de la transformation de R^N qui applique y sur ε à l'aide de l'équation (1.5.1) est triangulaire inférieure et les éléments diagonaux sont égaux à 1. La fonction de densité de y conditionnellement à $I_* = 0$ est donc donnée par (1.5.9) où ε_t est exprimé en fonction de y_t au moyen des relations (1.5.1). Nous aurons donc

$$\log L_C(\theta) = -N/2 \log(2\pi) - N/2 \log \sigma^2 - 1/2 \sum_{t=1}^N \varepsilon_t^2 \quad (1.5.10)$$

L'équation de vraisemblance correspondant à σ^2 fournit la relation suivante :

$$\sigma^2 = 1/N S_C(\phi, \theta) \quad (1.5.11)$$

par substitution dans (1.5.10), nous obtenons donc à une constante additive près

$$- N/2 \log(S_C(\phi, \theta)/N).$$

Ce qui montre que l'estimateur du maximum de vraisemblance de ϕ et de θ est tel que $S_C(\hat{\phi}, \hat{\theta})$ est minimum. Connaissant $\hat{\phi}$ et $\hat{\theta}$ nous déterminons l'estimateur $\hat{\sigma}^2$ de σ^2 par (1.5.11). Nous retrouverons ainsi le critère MCC du 5.2.1.

5.2.4. Evaluation de la fonction de vraisemblance non conditionnelle

Le critère du MV ne coïncide pas avec le critère MCN du 5.2.1 contrairement au cas conditionnel. il y a à cela une double raison :

1. \hat{I}_* devrait être l'espérance conditionnelle de I_* , étant donné y mais ce n'est pas le cas ici.
2. Le Jacobien de la transformation de ε à y n'est pas 1 mais dépend de ϕ et de θ . (Mélard (1985)).

La justification de la méthode MCN est donc faible, bien qu'elle soit souvent utilisée. A première vue l'évaluation de la fonction de vraisemblance n'est pas simple, nous pouvons en effet l'écrire sous la forme générale

$$L_V(\phi, \theta) = (2\pi)^{-N/2} \sigma^{-N} (\det \Gamma)^{-N/2} \exp\left\{-\frac{1}{2\sigma^2} y^T \Gamma^{-1} y\right\} \quad (1.5.12)$$

où Γ est la matrice de covariance du vecteur y et qui peut être obtenue à l'aide de (1.5.3).

Remarquons que l'équation de vraisemblance associée à σ^2 est

$$\sigma^2 = \frac{1}{N} y^T \Gamma^{-1} y \quad (1.5.13)$$

D'où $\log L_V(\phi, \theta)$ peut s'écrire, à une constante additive près, comme

$$-(N/2) \log[(1/N) y^T \Gamma y] - (1/2) \log(\det \Gamma) \quad (1.5.14)$$

Ce qui montre qu'il suffit de minimiser

$$S_V(\phi, \theta) = (\det \Gamma)^{1/N} (y^T \Gamma^{-1} y) \quad (1.5.15)$$

Il semble donc qu'il faut inverser une matrice $N \times N$ ce qui requiert un nombre d'opérations de l'ordre de N^3 multiplications et divisions et au moins N^2 mots en mémoire, ceci, même pour un ordinateur puissant, est relativement coûteux, d'autant plus qu'il est nécessaire d'utiliser des algorithmes d'optimisation, ce qui nécessite de nombreuses évaluations de la fonction de vraisemblance. Le choix de ce critère de comparaison entre différents algorithmes (nombre de multiplications et de divisions) est justifié par le fait que ce sont ces opérations qui prennent le plus de temps comparativement aux autres. Toutefois il est

montré (voir Zahaf (1998)) que parfois les résultats donnés par ce critère ne sont pas toujours vérifiés dans la pratique. Plusieurs algorithmes performants ont été développés à cette fin. En effet, de très nombreux travaux ont d'abord visé à développer un algorithme pour des cas très particuliers, par exemple : Siddiqui, (1958) pour les modèles $AR(p)$, Shaman, (1975) pour les modèles $MA(1)$ et $MA(q)$, Tiao et Ali, (1971) pour les modèles $ARMA(1,1)$. Par ailleurs d'autres travaux pour le cas plus général ont été élaborés depuis Newbold, (1974) puis Dent, (1977), Ali (1977) et Hall et Nicolls (1980). Ils ont permis de réduire nettement le temps de calcul. Ljung et Box (1979) ont pu synthétiser quelques-unes de ces méthodes et ont donné un algorithme qui réduit fortement la taille de la matrice à inverser. Ils utilisent des valeurs estimées et un algorithme d'inversion par blocs. Ansley (1979) a utilisé la factorisation de Cholesky de la matrice de covariance du processus étudié. En transformant le processus il obtient une matrice de covariance de type bande, c'est à dire dont seulement un nombre restreint de diagonales parallèles à la diagonale principale sont non nuls ce qui améliore fortement la rapidité de cet algorithme. Les méthodes de Ljung et box (1979) et d'Ansley (1979) requièrent un nombre d'opérations qui est borné par un polynôme de degré un. En se basant sur le filtre de Kalman (1960), Gardner, Harvey et Philips (1980) ont développé un algorithme qui permet d'avoir aussi un nombre d'opérations proportionnel à la taille de la série. Pearlman (1980) a proposé une méthode alternative qui consiste à remplacer le filtre de Kalman par d'autres relations de récurrence, qui porte le nom de Chandrasekhar et sont dues à More et *al.* (1974). Mélard (1984) a encore amélioré cet algorithme et l'a implanté. L'algorithme le plus rapide pour évaluer (1.5.12) dans le cas d'un modèle $ARMA(p,q)$ nécessite un nombre d'opérations d'ordre N . En pratique on évalue la fonction de vraisemblance exacte aussi rapidement que pour le cas des moindres carrés. Toutefois ces algorithmes sont trop longs pour pouvoir être exposés ici.

5.3 L'optimisation

Ayant défini le modèle paramétrique (il s'agit dans ce présent chapitre d'un modèle ARMA) et la fonction critère à optimiser (les critères de 5.1), nous allons construire des algorithmes de calcul permettant de réaliser cette tâche. Bien entendu la capacité de construire efficacement de tels algorithmes dépendra de façon importante de la complexité de la fonction critère et des propriétés de régularité. L'objet de ce paragraphe est de rappeler brièvement les deux grandes classes d'algorithmes numériques pour l'optimisation des critères retenus pour l'estimation des paramètres. Nous parlons d'optimisation déterministe dans la mesure où nous travaillons directement sur la fonction à optimiser qui sera supposée connue. Dans le cas contraire l'optimisation se fera à partir des données et l'algorithme sera nommé algorithme d'optimisation stochastique. Sauf dans des cas particuliers, nous devons minimiser une fonction qui est une forme quadratique des observations mais qui n'est pas quadratique en les paramètres. Généralement on ne peut même pas l'exprimer de façon compacte. Il faut donc recourir à un algorithme itératif de minimisation de fonction à plusieurs variables ou plus particulièrement (dans le cas d'un algorithme des moindres carrés) à un algorithme de minimisation d'une somme de carrés (algorithme de régression non linéaire). Il n'est donc pas possible de décrire ces algorithmes ici. Pour illustrer nous considérons simplement trois algorithmes usuels : l'algorithme du gradient, l'algorithme de Newton, et l'algorithme de Marquardt, pour une revue plus détaillée voir par exemple Harty, (1996) ou Kennedy et Gentle, (1980).

La fonction critère notée par $V(\beta)$ sera supposée au moins de classe C^2 sur le domaine D de β , et tous les algorithmes d'optimisation sont basés sur l'approximation locale de $V(\beta)$ par une fonction quadratique à l'aide de la formule de Taylor

$$V(\beta) = V(\beta_0) + \frac{\partial V(\beta_0)}{\partial \beta^T} [\beta - \beta_0] + \frac{1}{2} [\beta - \beta_0]^T \frac{\partial^2 V(\beta_0)}{\partial \beta \partial \beta^T} [\beta - \beta_0] + o(\|\beta - \beta_0\|^2)$$

d'où il résulte qu'une condition nécessaire d'optimum en β^* est la stationnarité

$$\frac{\partial V(\beta^*)}{\partial \beta^T} = 0 \quad \text{et} \quad \frac{\partial^2 V(\beta^*)}{\partial \beta \partial \beta^T} \geq 0 \quad (1.5.16)$$

c'est cette condition que nous cherchons à satisfaire par les algorithmes numériques.

Notons aussi que tous les algorithmes de calcul sont de la forme itérative suivante, construisant à partir d'un point initial β_0 (en général arbitraire), une suite d'approximations β_k d'une solution β^* de (1.5.16)

Etape 1 Choix de la valeur initiale β_0

Etape 2 Calcul itératif $\beta_{k+1} = \beta_k + \gamma_k S_k \frac{\partial V(\beta_k)}{\partial \beta^T}$

Etape 3 Critère d'arrêt : si le critère est vérifié alors fin, sinon aller en 2)

Avec : γ_k suite scalaire des pas de déplacement ou encore suite de gains scalaires.

S_k suite de matrices positives, gains matriciels de l'algorithme
suite de directions de déplacement.

Les algorithmes seront classés selon la méthode de choix des gains scalaires et des gains matriciels :

Nous parlons d'algorithme de gradient s'il n'y a pas de gain matriciel, c'est à dire : $S_k = I$.

Si la suite de gains matriciels approche, quand $k \rightarrow \infty$, l'inverse du hessien $\frac{\partial^2 V(\beta^*)}{\partial \beta \partial \beta^T}$ alors il s'agit d'un algorithme de Newton.

5.3.1. Algorithme du gradient

Cet algorithme s'applique à des situations générales (il n'est pas obligatoire que la fonction critère soit une somme de carrés). Prenons par exemple le critère MV avec le modèle (1.5.1) et notons par $\beta = (\phi^T, \theta^T)^T$ le vecteur des paramètres à estimer.

Sachant que

$$L_V(\beta) = L_V(\beta_0) + g(\beta^*)(\beta - \beta_0) \quad (1.5.16)$$

où

$$g(\beta^*) = \left. \frac{\partial L_V(\beta)}{\partial \beta^T} \right|_{\beta=\beta^*}$$

et chaque composante de β^* est comprise entre les composantes correspondantes de β et de β_0 , Nous minimisons $L_V(\beta)$ en prenant $g(\beta^*)(\beta - \beta_0)$ négatif, il suffit donc que chacune des composantes de $(\beta - \beta_0)$ soit de signe opposé à la composante correspondante du vecteur gradient de $L_V(\beta)$ en β^* . Comme β^* est inconnu nous prenons

$$\beta_1 = \beta_0 - k_1 g(\beta_0)^T \quad (1.5.17)$$

où $k_1 > 0$ est choisi de sorte que $L_V(\beta_1) < L_V(\beta_0)$. S'il s'avère difficile d'évaluer le gradient $g(\beta_0(i))$ au point $\beta_0(i)$ (où $\beta_0(i)$ est la $i^{\text{ème}}$ composante du vecteur β_0) nous le remplaçons par la différence divisée $\hat{g}(\beta_0(i))$ définie par :

$$\hat{g}(\beta_0(i)) = \frac{L_V(\beta_0(1), \dots, \beta_0(i) + k_2, \dots, \beta_0(p+q)) - L_V(\beta_0(1), \dots, \beta_0(i), \dots, \beta_0(p+q))}{k_2} \quad (1.5.18)$$

où k_2 est choisi selon l'instance du problème.

Ayant déterminé β_1 à partir de β_0 , nous itérons la procédure ce qui donne la suite $\beta_0, \beta_1, \dots, \beta_i, \beta_{i+1}, \dots$. Nous décidons d'un critère d'arrêt, par exemple

$$\frac{\|\beta_i - \beta_{i+1}\|}{\|\beta_i\|} < k_3 \quad (1.5.19)$$

Notons qu'il existe un lien entre la vitesse de convergence et le conditionnement du Hessien (voir Michaut, (1992)). Cette dépendance est une caractéristique de l'inconvénient majeur des méthodes de gradient, en effet la convergence de la méthode du gradient peut devenir extrêmement lente en cas de mauvais conditionnement du Hessien.

5.3.2. Algorithme de Gauss-Newton

Les méthodes de Gauss-Newton visent à éliminer l'inconvénient ci-dessus des méthodes du gradient, et à obtenir une vitesse de convergence meilleure que la première et indépendante du conditionnement du Hessien. Cette classe de méthodes utilise une approximation de l'ordre 1 de $\varepsilon_t(\beta)$. Choisissons par exemple le critère (MCC). Soit

$$\varepsilon_t(\beta) = \varepsilon_t(\beta_0) + f_t(\beta^*)(\beta - \beta_0) \quad (1.5.20)$$

où $f_t(\beta^*) = \left. \frac{\partial \varepsilon_t(\beta)}{\partial \beta^T} \right|_{\beta=\beta^*}$ est le vecteur gradient de ε_t au point β^* avec β^* défini comme ci

dessus. Donc nous pouvons minimiser la quantité $\sum_{t=1}^N (\varepsilon_t(\beta))^2$, où

$$\varepsilon_t(\beta_0) = -f_t(\beta_0)(\beta - \beta_0) + \varepsilon_t(\beta) \quad (1.5.21)$$

et $\varepsilon_t(\beta_0)$ est connu, ainsi que $f_t(\beta_0)$, approché par différence divisées comme ci-dessus. Par conséquent la méthode des moindres carrés ordinaire fournit

$$\beta - \beta_0 = -\frac{1}{2} g(\beta_0) \left[\sum_{t=1}^N f_t^T(\beta_0) f_t(\beta_0) \right]^{-1} \quad (1.5.22)$$

comme solution de (1.5.21), tout en remarquant que

$$\sum_{t=1}^N \varepsilon_t(\beta_0) f_t(\beta_0) = \frac{1}{2} g(\beta_0)$$

en pratique nous déterminons plutôt

$$\beta = \beta_0 - k_4 g(\beta_0) \left[\sum_{t=1}^N f_t^T(\beta_0) f_t(\beta_0) \right]^{-1} \quad (1.5.23)$$

où k_4 est tel que $S_c(\beta_1) < S_c(\beta_0)$.

Comme pour la méthode du gradient, nous itérons la procédure jusqu'à l'obtention de la solution à la précision choisie. Un inconvénient de la méthode de Newton est probablement la non garantie de descente et donc le risque de ne pas avoir une convergence globale.

Il existe cependant d'autres méthodes telles que les méthodes dites de compromis en outre l'algorithme de Marquardt.

5.3.3. Algorithme de compromis de Marquardt

L'algorithme de Marquardt qui s'applique uniquement pour l'estimation des moindres carrés combine de manière optimale la méthode de Gauss-Newton et la méthode du gradient. En effet, l'algorithme de Gauss-Newton, bien qu'il converge rapidement, il peut être absorbé par des extremums locaux. L'avantage de la méthode du gradient est qu'elle converge toujours vers les estimateurs des moindres carrés, mais tellement lentement qu'elle tend à devenir inefficace. Le compromis de Marquardt combine les avantages et non les défauts de chacune de ses méthodes. C'est-à-dire que celui-ci converge toujours très rapidement vers les estimateurs des moindres carrés. A partir de valeurs initiales données pour les coefficients, le compromis de Marquardt établit de nouveaux coefficients qui donnent une plus faible somme de carrés (RSS) et qui, de plus, sont très proches des coefficients à fonction critère minimum. L'algorithme progresse donc dans la bonne direction et, en plus, il choisit la correction optimale des coefficients c'est-à-dire qu'il arrive très rapidement vers la RSS minimale. A l'opposé de l'algorithme du gradients, le compromis de Marquardt ne se déplace pas d'une longueur constante dans la direction du gradient. La longueur du déplacement évalue en fonction de l'angle formé par les directions du gradient de deux itérations successives. Le compromis de Marquardt est une procédure itérative à huit étapes. Nous en proposons une illustration au travers de l'algorithme qui suit :

Etape 1 : l'algorithme choisit des valeurs initiales pour les $(p + q + 1)$ coefficients du modèle ARMA(p, q). Ces valeurs initiales sont placées dans un vecteur de taille $(p + q + 1)$.

Etape 2 : l'algorithme calcule la somme des carrés des résidus (RSS_0) associée à ces valeurs initiales. A ce niveau, la procédure est identique à la méthode de recherche par balayage.

Etape 3 : On calcule les dérivées de la méthode de Gauss-Newton.

Etape 4 : On détermine des approximations linéaires de la relation non linéaires entre les résidus et les coefficients estimés $\hat{\phi}_i, i = 1, \dots, p$ et $\hat{\theta}_j, j = 1, \dots, q$

Etape 5 : On établit un nouveau vecteur des coefficients ϕ_1 à l'aide d'un vecteur de corrections des équations linéarisées.

Etape 6 : Ces nouvelles estimations fournissent une nouvelle somme des carrés des résidus.

Etape 7 : on compare les sommes des carrés des résidus RSS_0 et RSS_1 pour savoir si l'algorithme converge vers un minimum.

Etape 8 : on calcule la variation relative de somme des carrés des résidus que l'on compare à un critère de convergence arbitraire α . Si cette variation relative est inférieure à α on peut supposer que le dernier vecteur des coefficients ϕ_1 est le vecteur des

paramètres des moindres carrés. Dans le cas contraire, on revient à l'étape 3 en ré-initialisant la RSS .

Ces huit étapes ne sont qu'une procédure de linéarisation de Gauss-Newton. Il est possible, qu'à l'étape 7, cette méthode ne tende pas vers une somme des carrés des résidus plus faible. C'est dans ce cas qu'intervient la méthode du gradient. Si, à l'étape 7, $RSS_1 > RSS_0$, on incrémente d'une certaine valeur un paramètre β défini au début du cycle, Les équations linéarisées, dont les paramètres dépendent de β , sont modifiées se qui conduit à nouvelles corrections Ainsi, à mesure que β augmente, les corrections apportées tendent vers les corrections optimales de la méthode du gradient.

6. Propriétés statistiques des estimateurs

Au 5.1, Nous avons envisagé plusieurs méthodes d'estimation, du point de vue asymptotique, c'est à dire quand la longueur N de la série tend vers l'infini, ces différentes méthodes sont équivalentes. Sous certaines conditions générales les estimateurs convergent (en probabilité) vers les vraies valeurs de paramètres et sont asymptotiquement distribuées selon la loi normale.

Une évaluation d'un estimateur n'est intéressante que si nous pouvons estimer la matrice de covariance associée. En effet les tests sur les paramètres, les méthodes de *scoring* et le problème de la détermination de la taille d'un échantillon nécessitent la connaissance de cette matrice. Par ailleurs, l'obtention de la matrice de covariance peut être faite à l'aide de l'inverse de la matrice d'information de Fisher. Nous commençons d'abord par définir la matrice d'information de Fisher pour les paramètres d'un modèle ARMA

6.1. Matrice d'information de Fisher

La matrice d'information de Fisher, F_N , des paramètres β d'un modèle ARMA, correspondant à la série d'observations y_1, y_2, \dots, y_N , est définie par

$$F_N = - \left[E \left(\frac{\partial^2 \log L(\beta)}{\partial \beta \partial \beta^T} \right) \right] \quad (1.6.1)$$

où $L(\cdot)$ est la fonction de vraisemblance de β relativement à la série y_1, y_2, \dots, y_N .

Whittle, (1953) a introduit la notion de matrice d'information asymptotique F en normalisant la matrice F_N par la taille de l'observation N . C'est à dire

$$F = \lim_{N \rightarrow \infty} N^{-1} F_N \quad (1.6.2)$$

Par ailleurs, la densité spectrale $f(\omega)$ d'un modèle ARMA est une fonction absolument continue de ω , donc $f(\omega)$ est strictement positive pour tout ω . Par suite les éléments de F peuvent être donnés par la formule de Whittle, (1953)

$$F_{jk} = \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{1}{f^2(\omega)} \frac{\partial f(\omega)}{\partial \beta_j} \frac{\partial f(\omega)}{\partial \beta_k} d\omega \quad 1 \leq j, k \leq p + q + 1 \quad (1.6.3)$$

où $f(\omega)$ est donnée par

$$f(\omega) = \sum_{h=-\infty}^{+\infty} \gamma_h \cos(h\omega) \quad (1.6.4)$$

avec γ_h est la fonction d'autocovariances du processus à l'horizon h .

6.2. Propriétés statistiques de l'estimateur du MV

Soit y_1, y_2, \dots, y_N une suite de v. a. indépendantes gaussiennes et identiquement distribuées de moyenne β et notons par $\hat{\beta}_N$ l'estimateur du maximum de vraisemblance de β alors

$\hat{\beta}_N$ est un estimateur convergent vers θ de plus

$\sqrt{N}[\hat{\beta}_N - \beta]$ est asymptotiquement une loi gaussienne de moyenne nulle et de variance l'inverse de la matrice d'information de Fisher.

Sous certaines conditions générales, les estimateurs convergent (en probabilité) vers les vraies valeurs des paramètres et ont une distribution limite la loi normale dont les caractéristiques sont à préciser. De ce fait la variance asymptotique de l'estimateur du maximum de vraisemblance peut être approchée par la matrice d'information de Fisher. Ainsi en considérant par exemple le critère MCC alors nous avons

$$\begin{aligned} V(\hat{\beta}) &= - \left[E \left(\frac{\partial^2 \log L_C(\beta)}{\partial \beta \partial \beta^T} \right) \right]^{-1} \\ &= 2\sigma^2 \left[E \left(\frac{\partial^2 S_C(\beta)}{\partial \beta \partial \beta^T} \right) \right]^{-1} \end{aligned} \quad (1.6.5)$$

or

$$E \left(\frac{\partial^2 S_C(\beta)}{\partial \beta \partial \beta^T} \right) = 2 \sum_{i=1}^N E \left[\frac{\partial \varepsilon_i(\beta)}{\partial \beta} \frac{\partial \varepsilon_i(\beta)}{\partial \beta^T} + \varepsilon_i(\beta) \frac{\partial^2 \varepsilon_i(\beta)}{\partial \beta \partial \beta^T} \right]$$

et $E \left[\varepsilon_i(\beta) \frac{\partial^2 \varepsilon_i(\beta)}{\partial \beta \partial \beta^T} \right] = 0$ car $\frac{\partial^2 \varepsilon_i(\beta)}{\partial \beta \partial \beta^T}$ s'écrit comme combinaison linéaire du passé de y_i

comme ε_i est l'innovation du processus il s'ensuit qu'elle est non corrélée avec le passé du processus et donc par toutes ses combinaisons linéaires.

Ainsi nous aurons :

$$V(\hat{\beta}) = \frac{\sigma^2}{N} \left[E \left(\frac{\partial \varepsilon_i(\beta)}{\partial \beta} \frac{\partial \varepsilon_i(\beta)}{\partial \beta^T} \right) \right]^{-1} \quad (1.6.6)$$

Comme nous allons voir plus loin, plusieurs algorithmes sont développés afin d'évaluer (1.6.6).

Chapitre 2

Processus Périodiquement corrélés et modèles ARMA périodiques

1. Introduction

L'introduction des modèles $ARIMA(p,d,q)$ par Box et Jenkins (1970) a permis la modélisation d'une gamme très étendue de séries chronologiques. En effet, il est montré que pour la classe des séries stationnaires, nous pouvons toujours trouver deux entiers p et q pour lesquels le modèle $ARMA(p,q)$ associé soit représentatif. Pour certaines séries non stationnaires il existe des transformations usuelles telles que la technique d'élimination de la tendance ou de la saisonnalité permettent de passer à des séries stationnaires.

Cependant, il existe des séries temporelles non stationnaires qui ne peuvent être rendues stationnaires au moyen de techniques citées ci-dessus. Et c'est dans ce besoin qu'on a élargit la classe des modèles ARMA à des modèles $ARMA_t(p_t, q_t)$, à paramètres dépendant du temps (les bases théoriques de cette construction sont assurées par la généralisation du théorème de Wold (1938) par Cramér (1961) au cas de processus non stationnaires). Une classe particulière des modèles $ARMA_t$ évolutifs dans le temps qui est souvent très utile pour la modélisation des séries à caractère saisonnier est la classe de modèles autorégressifs moyenne

mobile notée PARMA où les paramètres associés sont périodiques dans le temps. Du fait que la structure de la corrélation de certaines séries saisonnières dépend de la saison que l'application des techniques usuelles à ces séries ne permet pas de passer à des modèles adéquats : En effet sur l'exemple suivant (de Tiao et Grupe, (1980)) on aperçoit l'utilité de l'introduction d'un modèle périodique : Etant donné une série de mesures représentant la quantité d'ozone ambiant à chaque heure sur le bassin de Los-Angeles (cette série est saisonnière de saison 24), alors il est bien connu que la concentration de l'ozone dans ce bassin s'accumule typiquement durant les heures du matin, atteint le pic au début de l'après midi et se dissipe doucement pour atteindre le zéro vers la nuit. Donc pour un jour donné : Nous apercevons que la corrélation entre les valeurs de l'ozone des heures matinales et celles des heures de l'après midi soit nettement plus élevée que la corrélation entre les valeurs des heures matinales et celles des heures de la nuit précédente. Il est clair donc que l'application de la technique d'élimination de la saison ne rend pas la série obtenue stationnaire. Ce phénomène a été souvent observé dans la pratique et a été établi théoriquement par ces même auteurs et d'autres.

Dans un aspect purement théorique, Gladyshev, (1961) a introduit pour la première fois dans la littérature des processus aléatoires le concept des processus périodiquement corrélés. Ce sont des processus du second ordre dont la moyenne et la fonction d'autocovariance sont des fonctions périodiques du temps. Le profit principal tiré de cette nouvelle classe de processus est le lien, établi par ce même auteur, existant entre ces processus et les processus multivariés stationnaires qui leur sont associés. En effet, cet auteur a donné un théorème consistant à ce qu'une condition nécessaire et suffisante pour qu'un processus aléatoire soit périodiquement corrélé est que le processus multivarié qui lui correspond soit faiblement stationnaire. Par ailleurs, la généralisation, aux processus non stationnaires par Cramér (1961), du théorème de Wold (1938) qui est relative aux processus stationnaires du second ordre et qui consiste en une représentation linéaire de ces processus, a permis d'introduire les modèles linéaires à coefficients évolutifs dans le temps. Une classe particulière de ces modèles et qui permet de représenter les processus périodiquement corrélés est la classe des modèles linéaires à coefficients périodiques. Par la suite plusieurs travaux de recherches ont montré que les modèles linéaires périodiques PARMA (représentation linéaire des processus périodiquement corrélés) s'avèrent efficaces pour la modélisation des séries saisonnières.

Un autre intérêt des modèles périodiques, autre qu'ils permettent l'étude des phénomènes saisonniers, est qu'ils peuvent être exploités dans l'analyse des modèles ARMA multidimensionnelles, dans le but de réduire sensiblement le nombre de paramètres à prendre en considération (Pagano (1978)). En effet, l'introduction du théorème de Gladyshev (1961) a permis donc d'étendre et d'étudier les propriétés théoriques des modèles autorégressifs moyenne mobile périodiques. Ce résultat a donné naissance à une nouvelle technique dite 'period-span lumping' pour l'étude des modèles périodiques en exploitant les résultats existant dans la littérature des processus aléatoires multidimensionnels stationnaires. En effet l'étude des propriétés théoriques d'un modèle autorégressif moyenne mobile d -périodique (tels que la causalité et l'inversibilité,...) se ramène à celle du modèle autorégressif moyenne mobile d -varié stationnaire qui lui correspond.

2. Processus périodiquement corrélés

Dans ce qui suit nous allons emprunter de la littérature des processus aléatoires quelques principaux résultats concernant les processus périodiquement corrélés.

Soit $\{y_t, t \in \mathbb{Z}\}$ un processus du second ordre, de moyenne $\mu_t = E(y_t); t \in \mathbb{Z}$ et de fonction d'autocovariance

$$\gamma(t, s) = \text{cov}(y_t, y_s) \quad t, s \in \mathbb{Z}. \quad (2.2.1)$$

Nous supposons sans perdre de généralité que $\mu_t = 0, t \in \mathbb{Z}$ c'est à dire que $\{y_t, t \in \mathbb{Z}\}$ représente une déviation d'un processus $\{Z_t, t \in \mathbb{Z}\}$ de sa moyenne μ_t .

Définition 2.2.1 Processus périodique au sens strict

Le processus $\{y_t, t \in \mathbb{Z}\}$ est dit strictement périodique (ou périodique au sens fort) s'il existe un entier d positif tel que pour tout entier positif k et pour tout t_1, t_2, \dots, t_k , : la distribution conjointe de $(y_{t_1}, y_{t_2}, \dots, y_{t_k})^T$ est la même que celle de $(y_{t_1+d}, y_{t_2+d}, \dots, y_{t_k+d})^T$. Autrement dit, si :

$$F_{t_1+d, t_2+d, \dots, t_k+d}(x_1, x_2, \dots, x_k) = F_{t_1, t_2, \dots, t_k}(x_1, x_2, \dots, x_k) \quad (2.2.2)$$

où $F(\cdot)$ est la fonction de distribution du processus $\{y_t, t \in \mathbb{Z}\}$.

Le plus petit entier d positif, vérifiant la dernière condition, est appelé période du processus.

Remarque

Il est clair que le cas où $d = 1$ correspond au cas des processus stationnaires au sens strict.

Comme pour la stationnarité, il est également impossible en pratique d'étudier la périodicité stricte des processus aléatoires, sauf dans des cas très spéciaux. Toutefois, plusieurs propriétés importantes des processus aléatoires peuvent être obtenues juste à partir de leurs moments du premier et du second ordre. La périodicité dans ces deux moments peut être donc suffisante pour expliquer les propriétés périodiques du processus. D'où l'introduction (Gladyshev, 1961) d'une classe particulière des processus aléatoires appelés processus périodiquement corrélés.

Définition 2.2.2 Processus périodiquement corrélés

Le processus $\{y_t, t \in \mathbb{Z}\}$ est dit périodiquement corrélé (ou encore périodique au sens faible) s'il existe un entier d strictement positif, tel que :

$$\begin{aligned} i) \quad & \mu_{t+d\tau} = \mu_t \quad \forall t, \tau \in \mathbb{Z}. \\ ii) \quad & \gamma(t+d\tau, s+d\tau) = \gamma(t, s) \quad \forall s, t, \tau \in \mathbb{Z}. \end{aligned} \quad (2.2.3)$$

Autrement dit, si la moyenne et la fonction d'autocovariance sont d -périodiques dans le temps.

Remarque

Un processus faiblement stationnaire est un processus périodiquement corrélé 1-périodique ($d = 1$).

2.1. Fonction d'autocovariance d'un processus périodiquement corrélé

Soit $\{y_t, t \in \mathbb{Z}\}$ un processus périodiquement corrélé d -périodique de moyenne nulle et de fonction d'autocovariance $\gamma(\cdot, \cdot)$ d -périodique. Pour tout entier $t \in \mathbb{Z}$, $i = 1, \dots, d$ et $\tau \in \mathbb{Z}$, tel que : $t = i + d\tau$, alors la fonction d'autocovariance pour l'horizon h (ou d'ordre h) et relative à la $i^{\text{ème}}$ période (ou saison), notée $\gamma_h^{(i)}$, $i = 1, \dots, d$, est définie par :

$$\gamma_h^{(i)} = \gamma(t, t - h) = \gamma(i, i - h), \quad (2.2.4)$$

$i = 1, \dots, d, h \in \mathbb{N}$ et $t \in \mathbb{Z}$.

Il est facile de vérifier que :

$$\gamma_h^{(i+d\tau)} = \gamma_h^{(i)} \quad \text{et} \quad \gamma_{-h}^{(i)} = \gamma_h^{(i+h)}. \quad (2.2.5)$$

2.2. Fonction d'autocovariance empirique d'un processus périodiquement corrélé

Supposons que nous disposons d'une série d'observations $\{y_1, y_2, \dots, y_{Nd}\}$ générée à partir d'un processus périodiquement corrélé de moyenne nulle et de fonction d'autocovariance $\gamma_h^{(i)}$, alors la fonction d'autocovariance empirique pour la période i et à l'horizon h , notée $\tilde{\gamma}_h^{(i)}$, est définie par

$$\tilde{\gamma}_h^{(i)} = N^{-1} \sum_{\tau=0}^{N-1} y_{i+d\tau} y_{i+d\tau-h}. \quad (2.2.6)$$

où $y_{i+d\tau-l} = 0$ pour $i + d\tau - l < 1$ ou pour $i + d\tau - l > Nd$.

Lemme 2.2.1 (Convergence de la fonction d'autocovariance empirique)

Si $\{y_t, t \in \mathbb{Z}\}$ est un processus gaussien périodiquement corrélé. Alors la fonction d'autocovariance empirique $\tilde{\gamma}$, telle qu'elle a été définie dans (2.2.6), converge presque sûrement et en moyenne quadratique, quand $N \rightarrow +\infty$, vers la fonction d'autocovariance γ .

Preuve voir (Pagano, 1978) \square

Théorème 2.2.1 (Pagano (1978)) (Matrice d'information de Fisher pour un PAR)

Soit $\{y_t, t \in \mathbb{Z}\}$ un processus autorégressif d'ordre $p_t, t = 1, \dots, d$ d -périodique, notons par F la matrice d'information de Fisher relative aux paramètres du modèle et par $F(\alpha, \beta)$ l'élément de cette matrice correspondant aux paramètres α et β . Alors la matrice d'information de Fisher est diagonale par bloc et est donnée par

$$\begin{cases} F(\phi_{kj}, \phi_{ml}) = \delta_{km} \gamma(k-j, k-l) / \sigma_k^2 \\ F(\phi_{kj}, \sigma_m^2) = 0 \\ F(\sigma_k^2, \sigma_m^2) = \delta_{km} / 2\sigma_k^4 \end{cases} \quad (2.2.7)$$

pour $j = 1, \dots, p_k, l = 1, \dots, p_m, k, m = 1, \dots, d$.

Preuve voir (Pagano 1978) \square

2.3. Processus multivarié stationnaire associé à un processus univarié périodiquement corrélé

Lors de l'introduction, pour la première fois, de la notion de processus périodiquement corrélés, Gladyshev (1961) a montré que l'on peut toujours construire à partir d'un processus périodiquement corrélé d -périodique, le processus stationnaire d -varié qui lui est correspondant. Cette relation permet la représentation d'un processus d -périodique par un certain processus d -varié qui est stationnaire. Par conséquent l'étude des processus périodiques se ramène immédiatement à celle des processus stationnaires multivariés.

Théorème 2.2.2 (Gladyshev (1961))

Soit $\{y_t, t \in \mathbb{Z}\}$ un processus centré périodiquement corrélé de période d . Pour tout $t \in \mathbb{Z}$, l'entier $i \in \{0, 1, \dots, d-1\}$ et l'entier τ tel que : $t = i + d\tau$ on définit le processus d -varié $X(\tau)$ où : $X_i(\tau) = y_{i+d\tau}$. Ce processus est alors donné par le vecteur

$$\begin{aligned} X(\tau) &= (X_1(\tau), X_2(\tau), \dots, X_d(\tau))^T \\ &= (y_{1+d\tau}, y_{2+d\tau}, \dots, y_{d+d\tau})^T \end{aligned}$$

Alors $\{y_t, t \in \mathbb{Z}\}$ est périodiquement corrélé si et seulement si le processus multivarié correspondant est stationnaire.

2.4. Relation entre les autocovariances d'un processus multivarié stationnaire et celles d'un processus périodiquement corrélé

Soit $\Gamma(\tau_1, \tau_2)$ la matrice de variance-covariance de $X(\tau_1)$ et $X(\tau_2)$ définie par :

$$\Gamma(\tau_1, \tau_2) = (\Gamma_{ij}(\tau_1, \tau_2))_{i,j=1,\dots,d} \quad \text{où} \quad \Gamma_{ij}(\tau_1, \tau_2) = \text{cov}(X_i(\tau_1), X_j(\tau_2)) \quad i, j = 1, \dots, d \text{ et } \tau_1, \tau_2 \in \mathbb{Z},$$

nous avons alors

$$\Gamma_{ij}(\tau_1, \tau_2) = \gamma_{i-j+d(\tau_1-\tau_2)}^{(i)} \quad i, j = 1, \dots, d \quad (2.2.8)$$

Le processus d -varié $X(\tau)$ étant stationnaire, alors sa matrice d'autocovariance à l'horizon h est de la forme :

$$\Gamma(h) = \Gamma(\tau, \tau - h) = (\Gamma_{ij}(k))_{i,j=1,\dots,d}$$

où $\Gamma_{ij}(k) = \gamma_{i-j+dk}^{(i)}$.

2.5. Processus bruit blanc périodique

Les processus bruits blancs stationnaires sont souvent utilisés pour générer d'autres processus stationnaires. De manière analogue, les processus bruits blancs périodiques, quant à eux jouent le même rôle dans le cas de processus périodiquement corrélés. Ainsi le processus $\{\varepsilon_t, t \in \mathbb{Z}\}$ est dit processus bruit blanc périodique s'il vérifie les propriétés suivantes :

- i) $\{\varepsilon_t, t \in \mathbb{Z}\}$ est centré, c'est à dire : $E(\varepsilon_t) = 0$ pour tout $t \in \mathbb{Z}$.
- ii) Il est de variance d -périodique : $\sigma_{i+d\tau}^2 = \sigma_i^2$ pour tout $i = 1, \dots, d$ et $\tau \in \mathbb{Z}$.
- iii) Le processus $\{\varepsilon_t, t \in \mathbb{Z}\}$ est non corrélé : $E(\varepsilon_t \varepsilon_s) = 0$, pour tout t et s tel que $t \neq s$.

3. Prédiction des processus périodiquement corrélés

Supposons que nous disposons de r variables aléatoires *connues* et que nous désirons d'avoir, sur la base de celles-ci, une approximation d'une certaine variable aléatoire *inconnue*. Alors une bonne approximation de celle-ci sera la combinaison linéaire des r variables dont les coefficients seront choisis de manière optimale. Dans ce qui suit nous allons étudier ce genre d'approximations et ce, pour les variables aléatoires d'un processus périodiquement corrélé.

3.1. Equation de prédiction

Soit $L^2(\Omega, \mathcal{A}, \mathbb{P})$ (ou simplement L^2) l'espace de Hilbert des variables aléatoires de carrés intégrables, muni du produit scalaire $\langle \cdot, \cdot \rangle$ et défini par : $\langle X, Y \rangle = E(XY)$. Et considérons le processus du second ordre $\{y_t, t \in \mathbb{Z}\}$. Notons aussi par $M_{t,r}, t \in \mathbb{Z}, r \in \mathbb{N} - \{0\}$, le sous-espace de Hilbert engendré par les r variables aléatoires : $y_{t-r}, y_{t-r+1}, \dots, y_{t-1}$. Alors le prédicteur linéaire (ou le meilleur prédicteur linéaire au sens des moindres carrés) $\hat{y}_t^{(r)}$ de y_t basé sur $M_{t,r}$ est défini comme la projection orthogonale (au sens de L^2) de y_t dans le sous-espace $M_{t,r}$, c'est à dire :

$$\hat{y}_t = P_{M_{t,r}} y_t \quad (2.3.1)$$

3.1.1. Equation de prédiction prospective

Puisque $\hat{y}_t^{(r)} \in M_{t,r}$ donc $\hat{y}_t^{(r)}$ peut s'écrire comme combinaison linéaire des $y_{t-r}, y_{t-r+1}, \dots, y_{t-1}$. Ainsi il existe des coefficients $\phi_{ij}^{(r)}$, $j = 1, \dots, r$ tels que :

$$\hat{y}_t^{(r)} = \sum_{j=1}^r \phi_{ij}^{(r)} y_{t-j} \quad (2.3.2)$$

où les coefficients $\phi_{ij}^{(r)}$, $j = 1, \dots, r$ satisfont les équations de prédiction, c'est à dire que :

$$\langle y_t - \hat{y}_t^{(r)}, y_{t-j} \rangle = 0 \quad \text{pour } j = 1, \dots, r$$

ou encore

$$\sum_{k=1}^r \phi_{ik}^{(r)} \langle y_{t-k}, y_{t-j} \rangle = \langle y_t, y_{t-j} \rangle \quad j = 1, \dots, r$$

d'où

$$\sum_{k=1}^r \phi_{ik}^{(r)} \gamma(t-k, t-j) = \gamma(t, t-j), \quad j = 1, \dots, r \quad (2.3.3)$$

où $\gamma(.,.)$ est la fonction d'autocovariance du processus $\{y_t, t \in Z\}$. La dernière équation peut se mettre également sous une forme matricielle,

$$\Gamma_{t,r} \Phi_{t,r} = \gamma_{t,r}$$

où $\Phi_{t,r} = (\phi_{t1}^{(r)}, \phi_{t2}^{(r)}, \dots, \phi_{tr}^{(r)})^T$, $\gamma_{t,r} = (\gamma(t, t-1), \gamma(t, t-2), \dots, \gamma(t, t-r))^T$ et $\Gamma_{t,r} = (\Gamma_{t,r})_{jk}$
 $\Gamma_{t,r}(j, k) = \gamma(t-k, t-j)$, $j, k = 1, \dots, r$

Si de plus la matrice $\Gamma_{t,r}$ est non singulière alors les coefficients du meilleur prédicteur linéaire $\hat{y}_t^{(r)}$ de y_t peuvent être trouvés au moyen de l'équation suivante

$$\Phi_{t,r} = \Gamma_{t,r}^{-1} \gamma_{t,r} \quad (2.3.4)$$

3.1.2. Equation de prédiction rétrospective

Notons maintenant par $\tilde{y}_t^{(r)}$ la meilleure projection (au sens des moindres carrés) de y_t sur le sous espace $M_{t+r+1,r}$ engendré par les r variables $y_{t+1}, y_{t+2}, \dots, y_{t+r}$. Comme $\tilde{y}_t^{(r)}$ est dans l'espace $M_{t+r+1,r}$ donc il existe des coefficients $\beta_{ij}^{(r)}$, $j = 1, \dots, r$ tels que :

$$\tilde{y}_t^{(r)} = \sum_{j=1}^r \beta_{ij}^{(r)} y_{t+j} \quad (2.3.5)$$

Puisque $\hat{y}_t \in M_{t,r}$ donc \hat{y}_t peut s'écrire comme combinaison linéaire des $y_{t-r}, y_{t-r+1}, \dots, y_{t-1}$.
où les coefficients $\beta_{ij}^{(r)}$, $j = 1, \dots, r$ satisfont les équations de prédiction, c'est à dire que :

$$\langle y_t - \tilde{y}_t^{(r)}, y_{t+j} \rangle = 0 \quad \text{pour } j = 1, \dots, r$$

ou encore

$$\sum_{k=1}^r \beta_{ik}^{(r)} \langle y_{t+k}, y_{t+j} \rangle = \langle y_t, y_{t+j} \rangle \quad j = 1, \dots, r$$

d'où

$$\sum_{k=1}^r \beta_{ik}^{(r)} \gamma(t+k, t+j) = \gamma(t, t+j), \quad j = 1, \dots, r \quad (2.3.6)$$

où $\gamma(.,.)$ est la fonction d'autocovariance du processus $\{y_t, t \in \mathbb{Z}\}$.

Remarques

- Si le processus $\{y_t, t \in \mathbb{Z}\}$ est stationnaire alors pour tout $t \in \mathbb{Z}$, la matrice $\Gamma_{t,r}$ et le vecteur $\gamma_{t,r}$ sont indépendants de t c'est à dire que pour tout t nous aurons : $\Gamma_{t,r} = \Gamma_r$ et $\gamma_{t,r} = \gamma_r$. Par suite les coefficients $\phi_{ij}^{(r)}$, $j = 1, \dots, r$ eux aussi ne dépendent que de r . Ainsi $\Phi_{t,r} = \Phi_r$, donc nous pouvons réécrire (2.3.4) comme suit :

$$\Phi_r = \Gamma_r^{-1} \gamma_r \quad (2.3.7)$$

- Si le processus $\{y_t, t \in \mathbb{Z}\}$ est périodiquement corrélé de période d , alors la fonction d'autocovariance est d -périodique, par suite la matrice $\Gamma_{t,r}$ et le vecteur $\gamma_{t,r}$ sont périodiques, donc nous aurons pour tout $t, \tau \in \mathbb{Z}$ et $i = 1, \dots, d$ tel que $t = i + d\tau$:

$$\Gamma_{i+d\tau,r} = \Gamma_{i,r} \quad \text{et} \quad \gamma_{i+d\tau,r} = \gamma_{i,r} \quad i = 1, \dots, d, \tau \in \mathbb{Z}.$$

si de plus les matrices $\Gamma_{i,r}$ pour $i = 1, \dots, d$ sont non singulières alors les coefficients de prédiction seront d -périodiques. D'où :

$$\phi_{i+d\tau,j}^{(r)} = \phi_{ij}^{(r)} \quad i = 1, \dots, d, \tau \in \mathbb{Z}, j = 1, \dots, r. \quad (2.3.8)$$

L'équation (2.3.3) peut s'écrire également comme :

$$\sum_{k=1}^r \phi_{ik}^{(r)} \gamma_{j-k}^{(i-k)} = \gamma_j^{(i)}, \quad j = 1, \dots, r \quad i = 1, \dots, d \quad (2.3.9)$$

Maintenant nous allons donner, pour le cas d'un processus stationnaire, une condition suffisante pour que la matrice Γ_r soit non singulière.

Proposition 2.3.1

Si la fonction d'autocovariance γ d'un processus $\{y_t, t \in Z\}$ stationnaire vérifie les conditions suivantes : $\gamma(0) > 0$ et $\gamma(h) \xrightarrow{h \rightarrow \infty} 0$ alors la matrice $\Gamma_r = (\gamma(i-j))_{i,j=1,\dots,r}$ de covariance du vecteur $(y_{t-r}, y_{t-r+1}, \dots, y_{t-1})^T$ est non singulière pour tout t et r . (Pour la preuve de la proposition le lecteur est renvoyé chez Brockwell et Davis, (1986)).

4. Modèle autorégressif moyenne mobile périodique

Avec l'incapacité des modèles SARIMA, traités dans le fameux ouvrage de Box et Jenkins (1970), à décrire et à représenter certaines séries à caractère non régulier (des séries générées par des processus non stationnaires) mais aussi avec la parution de la généralisation, par Cramér (1961), du théorème de Wold (1938), de la décomposition des processus stationnaires aux processus non stationnaires, les modèles linéaires non stationnaires à coefficient dépendants du temps (ARMA_t, évolutifs) ont connu un développement très important ces dernières années et sont souvent utilisés malgré leur complexité par rapport aux modèles linéaires stationnaires. Par analogie aux processus stationnaires et aux modèles ARMA qui les représentent, une classe particulière de modèles non stationnaires ARMA_t, évolutifs et qui permet de représenter les processus périodiquement corrélés est la classe de modèles linéaires à coefficients périodiques dans le temps. Ce paragraphe est consacré à l'étude de ce type de modèles. Nous commençons d'abord par présenter le théorème de Wold-Cramér pour la décomposition des processus non stationnaires.

4.1. Représentation linéaire des processus périodiquement corrélés

Du fait que les processus périodiquement corrélés sont des processus non stationnaires, au sens faible, la représentation de ces processus par la famille des modèles classiques ARMA n'est pas justifiée par le théorème de Wold (1938) qui assure l'existence et l'unicité d'une décomposition linéaire uniquement pour les processus, du second ordre, faiblement stationnaires. Néanmoins, les modèles linéaires autorégressifs moyenne mobile intégrés saisonniers étudiés par Box et Jenkins (1976) ont montré leur efficacité pour modéliser (après les avoir ramener au cas stationnaire à l'aide d'une ou de plusieurs transformations) plusieurs séries saisonnières et sont justifiés par la représentation de Wold. Toutefois, cette classe de modèles ne s'avère adéquate que pour une classe particulière de processus périodiquement corrélés (saisonniers) à savoir ceux qui peuvent être ramenés par le biais d'une transformation saisonnière ou mixte au cas stationnaire faible et par conséquent peuvent être modélisés par des membres de la famille des modèles saisonniers de Box et Jenkins. La pratique a montré qu'une diversité de processus périodiquement corrélés ne peuvent pas être modélisés de façon adéquate par cette classe de modèles intégrés saisonniers. Le théorème suivant de Cramér (1961) qui a donné une impulsion remarquable à l'analyse des séries chronologiques non stationnaires a permis l'introduction des modèles linéaires évolutifs, dans le temps, en particulier, les modèles ARMA périodiques.

4.2. Représentation linéaire de Wold-Cramér

Une généralisation du cas d'un processus, du second ordre, non stationnaire, du théorème de Wold (1938) concernant l'existence et l'unicité de la décomposition linéaire d'un

processus du second ordre faiblement stationnaire, a été établie par Cramér (1961). Ce résultat, connu sous le l'appellation décomposition (linéaire) de Wold-Cramér, a permis l'extension des modèles linéaires usuels (à coefficients constants) ARMA(p, q) aux modèles linéaires non-stationnaires, à coefficients dépendants du temps, ARMA(p_t, q_t), dont les modèles autorégressifs moyennes mobiles périodiques sont des cas particuliers.

4.3. Processus purement indéterminable et processus déterminable du second ordre

Notons par $H^2(y_t, t)$ le sous-espace de Hilbert, $H^2(y_t,]-\infty, t])$, engendré par le passé et le présent $(\dots, y_{t-2}, y_{t-1}, y_t)$ du processus du second ordre $\{y_t, t \in \mathbb{Z}\}$.

Définition 2.4.1

Un processus stochastique, du second ordre, $\{y_t, t \in \mathbb{Z}\}$ est dit purement indéterminable ou encore régulier, si

$$\lim_{t \rightarrow -\infty} H^2(y_t, t) = \{0\} \quad (2.4.1)$$

et purement déterminable ou singulier si

$$\lim_{t \rightarrow -\infty} H^2(y_t, t) = \lim_{t \rightarrow +\infty} H^2(y_t, t) \quad (2.4.2)$$

Théorème 2.4.1 (Cramér (1961))

Tous processus stochastique, du second ordre, $\{y_t\}$ possède une décomposition linéaire unique donnée par :

$$y_t = U_t + V_t \quad (2.4.3)$$

telle que :

- Les processus U_t et V_t appartiennent au sous-espace de Hilbert $H^2(y_t, t)$.
- Ces processus sont orthogonaux, de plus le processus U_t est purement déterminable (singulier) et V_t est purement indéterminable (régulier) et peut être représenté par une combinaison linéaire infinie unique convergente (en moyenne quadratique) de la forme

$$V_t = \sum_{s=-\infty}^t a_{t,s} \varepsilon_s \quad (2.4.4)$$

où $\{\varepsilon_t\}$ est un processus bruit blanc. Le processus $\{a_{t,t} \varepsilon_t, t \in \mathbb{Z}\}$ est, alors, dit innovation du processus et le processus $\{\varepsilon_t\}$ est l'innovation normée.

Cette décomposition a permis donc la construction des modèles linéaires non stationnaires qui ne sont que des approximations de la représentation de Wold-Cramér. Dans ce qui suit nous présenterons une classe particulière de ces modèles qui est la classe des modèles autorégressifs moyennes mobiles périodiques.

Considérons un processus périodiquement corrélé généré à partir d'un bruit blanc périodique $\{\varepsilon_t, t \in Z\}$

$$y_t = \varepsilon_t + \sum_{s=1}^{\infty} a_{t,s} \varepsilon_{t-s} \quad (2.4.5)$$

où la suite $\{a_{t,s}\}$ est absolument sommable pour tout t , alors les coefficients $\{a_{t,s}\}$ sont périodiques dans le temps (en t).

4.4. Modèle ARMA périodique (PARMA)

Comme pour les modèles ARMA classiques, un processus périodiquement corrélé peut être représenté par un modèle ARMA périodique qui n'est rien d'autre qu'une approximation de la décomposition de Wold-Cramér.

Définition 2.4.2 (Processus Autorégressif moyenne mobile périodique)

Le processus périodiquement corrélé $\{y_t, t \in Z\}$ d -périodique est dit admettre une représentation autorégressive moyenne mobile d'ordre (p_t, q_t) , périodique de période d , noté PARMA $_d(p_t, q_t)$, s'il est solution de l'équation aux différences stochastique suivante :

$$y_t - \sum_{j=1}^{p_t} \phi_{tj} y_{t-j} = \varepsilon_t - \sum_{j=1}^{q_t} \theta_{tj} \varepsilon_{t-j} \quad (2.4.6)$$

ou encore :

$$\Phi_t(L)X_t = \Theta_t(L)\varepsilon_t \quad t \in Z$$

où : $\{\varepsilon_t, t \in Z\}$ est un bruit blanc périodique de variance σ_t^2 les paramètres θ_{tj} $j = 1, \dots, q_t$; ϕ_{ti} $i = 1, \dots, p_t$; q_t et p_t sont des fonctions d -périodiques et les polynômes $\Phi_t(B) = 1 - \sum_{i=1}^{p_t} \phi_{ti} B^i$ et $\Theta_t(B) = 1 - \sum_{j=1}^{q_t} \theta_{tj} B^j$ ne possèdent pas de zéros communs.

Posons $p = \max_{1 \leq t \leq d} p_t$ et $q = \max_{1 \leq t \leq d} q_t$ alors (2.4.6) peut s'écrire également dans une forme plus simple

$$y_t - \sum_{j=1}^p \phi_{tj} y_{t-j} = \varepsilon_t - \sum_{j=1}^q \theta_{tj} \varepsilon_{t-j} \quad (2.4.7)$$

4.5. Modèle autorégressif périodique (PAR)

Définition 2.4.3. (Processus Autorégressif périodique)

Le processus périodiquement corrélé $\{y_t, t \in Z\}$ d -périodique est dit admettre une représentation autorégressive périodique d'ordre p_t , de période d , noté $\text{PAR}_d(p_t)$, s'il est solution de l'équation aux différences stochastique suivante :

$$y_t - \sum_{i=1}^{p_t} \phi_{ti} y_{t-i} = \varepsilon_t \quad t \in Z; \quad (2.4.8)$$

où : $\{\varepsilon_t, t \in Z\}$ est un bruit blanc périodique de variance σ_t^2 , les paramètres $\phi_{ti} \quad i=1, \dots, p_t$; et p_t et la variance σ_t^2 sont des fonctions d -périodiques.

Le modèle $\text{PAR}_d(p_t)$ est un modèle particulier, du modèle d -périodique général correspondant au cas où $\theta_{ij} = 0$ pour $j=1, 2, \dots, q_t$, et $t \in Z$.

4.5.1. Equations de Yule-Walker périodiques

Comme pour le cas des modèles autorégressifs univariés classiques, les paramètres $\phi_{il}, \quad i=1, \dots, d; l=1, \dots, p_i$ et $\sigma_j; \quad j=1, \dots, d$ du processus périodique $\{y_t, t \in Z\}$ sont liés avec la fonction d'autocovariance au moyen des équations de Yule-Walker dans une forme modifiée appelée souvent les équations de Yule-Walker périodiques.

Théorème 2.4.2 (Les équations de Yule-Walker périodiques)

Si $\{y_t, t \in Z\}$ est un processus autorégressif d -périodique d'ordre p_t de fonction d'autocovariance γ alors pour tout $k=1, \dots, d$ Nous avons :

$$\gamma(k, k-v) - \sum_{j=1}^{p_k} \phi_{kj} \gamma(k-j, k-v) = \delta_{v0} \sigma_k^2 \quad v \geq 0. \quad (2.4.9)$$

où δ_{v0} désigne le symbole *delta Kronecker* (défini par $\delta_{v0} = \begin{cases} 1 & \text{si } v > 0 \\ 0 & \text{sinon} \end{cases}$)

Preuve :

Le processus $\{y_t, t \in Z\}$ étant autorégressif d -périodique d'ordre $p_t \quad t=1, \dots, d$, , alors pour tout $t \in Z$ posons $t = k + d\tau$ avec $k=1, \dots, d$ et $\tau \in Z$. Ainsi Nous avons

$$y_t - \sum_{j=1}^{p_t} \phi_{tj} y_{t-j} = \varepsilon_t,$$

qui peut s'écrire autrement

$$y_{k+d\tau} - \sum_{j=1}^{p_k} \phi_{k+d\tau, j} y_{k+d\tau-j} = \varepsilon_{k+d\tau} \quad , k=1, \dots, d. \quad (2.4.10)$$

Multiplions les membres de l'équation (2.4.10) par y_{t-v} (ou également par $y_{k+d\tau-v}, k=1, \dots, d$). Pour v positif ε_t est supposée non corrélée avec y_{t-v} en prenons

donc l'espérance mathématique des deux membres de l'équation nous obtenons le résultat du théorème (voir aussi Pagano, 1978). \square

4.5.2. Matrice d'information de Fisher limite d'un modèle PAR

La matrice d'information de Fisher asymptotique des paramètres d'un modèle ARMA a déjà été définie au chapitre 1. Pour un modèle autorégressif périodique la définition est la même, notons seulement que la relation entre la matrice limite est la matrice exacte dans le cas périodique n'est pas la même que celle dans le cas classique, en effet pour un modèle ARMA périodique, la matrice d'information de Fisher asymptotique sera normalisée par la taille du canal et non pas par la taille de l'observation.

Soit F_{Nd} la matrice d'information de Fisher pour les paramètres d'un modèle PARMA relativement à une série d'observation y_1, y_2, \dots, y_{Nd} alors nous avons la relation suivante

$$F = \lim_{N \rightarrow \infty} (1/N) F_{Nd}$$

Théorème 2.4.3 (Matrice d'information de Fisher pour un AR périodique)

Soit $\{y_t, t \in Z\}$ un processus autorégressif d'ordre $p_t, t = 1, \dots, d$ d -périodique, notons par F la matrice d'information de Fisher relative aux paramètres du modèle et par $F(\alpha, \beta)$ l'élément de cette matrice correspondant aux paramètres α et β . Alors la matrice d'information de Fisher est diagonale par bloc et est donnée par

$$\begin{cases} F(\phi_{kj}, \phi_{ml}) = \delta_{km} \gamma(k-j, k-l) / \sigma_k^2 \\ F(\phi_{kj}, \sigma_m^2) = 0 \\ F(\sigma_k^2, \sigma_m^2) = \delta_{km} / 2\sigma_k^4 \end{cases} \quad (2.4.11)$$

pour $j = 1, \dots, p_k, l = 1, \dots, p_m, k, m = 1, \dots, d$.

Preuve voir (Pagano, 1978) \square

4.6. Modèles moyenne mobile périodique (PMA)

Définition 2.4.4. (processus moyenne mobile périodique)

Le processus périodiquement corrélé $\{y_t, t \in Z\}$ d -périodique est dit admettre une représentation moyenne mobile périodique d'ordre q_t de période d , noté MA(q_t), s'il est solution de l'équation aux différences stochastique suivante :

$$y_t = \varepsilon_t - \sum_{j=1}^{q_t} \theta_{ij} \varepsilon_{t-j} \quad t \in Z; \quad (2.4.12)$$

où les paramètres $\theta_{ij} \quad j=1, \dots, q_t; t \in Z$ et la variance du bruit blanc σ_t^2 sont des fonctions périodiques, en t , de période d .

Le modèle MA(q_t) est un modèle particulier, du modèle d -périodique général correspondant au cas où $\phi_{ij} = 0$, pour $j=1, 2, \dots, q_t$, et $t \in Z$.

5. Méthode d'études des modèles ARMA périodiques

Avant de passer à l'étude des modèles ARMA périodiques nous commençons d'abord par rappeler deux propriétés essentielles qui sont la causalité et l'inversibilité.

Définition 2.5.1. Causalité

Le modèle (2.21) est dit causal s'il admet une solution de la forme

$$y_t = \sum_{i=0}^{\infty} \alpha_i \varepsilon_{t-i} \quad t \in Z \quad (2.5.1)$$

où la série infinie est convergente, en moyenne quadratique. L'expression précédente est dite représentation de Wold-Cramer du processus $\{y_t, t \in Z\}$.

Définition 2.5.2. Inversibilité

Le modèle (2.21) est dit inversible s'il existe des fonctions β_i telles que :

$$\varepsilon_t = \sum_{i=0}^{\infty} \beta_i y_{t-i} \quad t \in Z \quad (2.5.2)$$

où la série est convergente, en moyenne quadratique. L'expression, convergente, précédente est dit formule de déconvolution.

Pour étudier les propriétés des modèles ARMA périodiques nous pouvons distinguer deux approches. La première se base directement sur le théorème de Gladyshev (1961). Et consiste à ramener un modèle périodique univarié à un modèle classique multivarié et d'étudier les propriétés de ce dernier. Quant à la deuxième, elle ramène un modèle moyenne mobile (autorégressif) périodique à un modèle moyenne mobile périodique (autorégressif) mais seulement d'ordre 1.

5.1. Technique « period-span lumping » et étude des modèles univariés ARMA_d périodiques

La technique la plus ancienne et la plus couramment utilisée pour étudier diverses propriétés théoriques des modèles autorégressifs moyennes mobiles périodiques univariés est la technique dite "Period-Span Lumping". L'essence de cette méthode est de ramener l'étude des propriétés d'un modèle périodique univarié à celle du modèle d -varié autorégressif moyenne mobile qui lui correspond. Ceci permet d'exploiter les résultats déjà existant concernant l'analyse des modèles multivariés classiques pour étudier les modèles périodiques. Dans ce qui suit nous allons présenter le lien entre un modèle périodique univarié et un modèle multivarié classique.

5.2. Modèle ARMA multivarié associé à un modèle ARMA_d(p_t,q_t) univarié

Soit $\{y_t, t \in \mathbb{Z}\}$ un processus univarié périodiquement corrélé admettant une représentation autorégressive moyenne mobile d'ordre (p_t, q_t) , périodique de période d

$$y_t - \sum_{j=1}^{p_t} \phi_{tj} y_{t-j} = \varepsilon_t - \sum_{j=1}^{q_t} \theta_{tj} \varepsilon_{t-j}. \quad (2.5.3)$$

Alors le processus d -varié $\{X(\tau)\}$ définit par : $X_i(\tau) = y_{i+d\tau}$

où

$$X(\tau) = (X_1(\tau), X_2(\tau), \dots, X_d(\tau))^T \\ = (y_{1+d\tau}, y_{2+d\tau}, \dots, y_{d+d\tau})^T$$

satisfait à un modèle d -varié autorégressif moyenne mobile donné ci-dessous

$$\phi(0)X(\tau) + \phi(1)X(\tau-1) + \dots + \phi(p^*)X(\tau-p^*) = \vartheta(0)\eta(\tau) + \theta(1)\eta(\tau-1) + \dots + \theta(q^*)\eta(\tau-q^*) \quad (2.5.4)$$

où

$\eta(\tau) = (\eta_1(\tau), \eta_2(\tau), \dots, \eta_s(\tau))^T$ avec $\eta_i(\tau) = \varepsilon_{i+d\tau}$, $i = 1, \dots, d$ est un processus bruit blanc d -varié.

$$p^* = \max_{1 \leq i \leq d} \left\lceil \frac{p_i - i}{d} \right\rceil + 1 \quad \text{et} \quad q^* = \max_{1 \leq i \leq d} \left\lceil \frac{q_i - i}{d} \right\rceil + 1;$$

$$\Phi(0) = \begin{pmatrix} 1 & 0 & \cdot & \cdot & \cdot \\ \phi_1(2) & 1 & 0 & \cdot & \cdot \\ \cdot & \cdot & 1 & 0 & \cdot \\ \phi_{i-1}(i) & \cdot & \cdot & 1 & \cdot \\ \phi_{d-1}(d) & \cdot & \cdot & \phi_1(d) & 1 \end{pmatrix} \quad \text{et} \quad \Theta(0) = \begin{pmatrix} 1 & 0 & \cdot & \cdot & \cdot \\ \theta_1(2) & 1 & 0 & \cdot & \cdot \\ \cdot & \cdot & 1 & 0 & \cdot \\ \theta_{i-1}(i) & \cdot & \cdot & 1 & \cdot \\ \theta_{d-1}(d) & \cdot & \cdot & \theta_1(d) & 1 \end{pmatrix} \quad (2.5.5)$$

et

$$\Phi(k) = \begin{pmatrix} \phi_{dk}(1) & \phi_{dk-1}(1) & \phi_{dk-2}(1) & \cdot & \cdot & \cdot & \cdot & \cdot & \phi_{dk+1-d}(1) \\ \phi_{dk+1}(2) & \phi_{dk}(2) & \phi_{dk-1}(2) & \cdot & \cdot & \cdot & \cdot & \cdot & \phi_{dk+2-d}(2) \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \phi_{dk+i-1}(i) & \cdot & \cdot & \cdot & \phi_{dk+1}(i) & \phi_{dk}(i) & \phi_{dk-1}(i) & \cdot & \phi_{dk+i-d}(i) \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \phi_{dk+d-1}(d) & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \phi_{dk+1}(d) & \phi_{dk}(d) \end{pmatrix} \quad (2.5.6)$$

avec $i=1,2,\dots,d$ et $k=1,2,\dots,p$

et

$$\Theta(r) = \begin{pmatrix} \theta_{dr}(1) & \theta_{dr-1}(1) & \theta_{dr-2}(1) & \cdot & \cdot & \cdot & \cdot & \cdot & \theta_{dr+1-d}(1) \\ \theta_{dr+1}(2) & \phi_{dr}(2) & \theta_{dr-1}(2) & \cdot & \cdot & \cdot & \cdot & \cdot & \theta_{dr+2-d}(2) \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \theta_{dr+i-1}(i) & \cdot & \cdot & \cdot & \theta_{dr+1}(i) & \theta_{dr}(i) & \theta_{dr-1}(i) & \cdot & \theta_{dr+i-d}(i) \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \theta_{dr+d-1}(d) & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \theta_{dr+1}(d) & \theta_{dr}(d) \end{pmatrix} \quad (2.5.7)$$

avec $i=1,2,\dots,d$ et $r=1,2,\dots,q$.

Cette relation entre un $\text{ARMA}_d(p_t, q_t)$ univarié et un $\text{ARMA}(p^*, q^*)$ d -varié permet de vérifier quelques propriétés théoriques tel que la causalité et l'invisibilité.

Le modèle d -varié autorégressif moyenne mobile (2.5.4) associé au modèle périodique (2.5.3) ainsi obtenu peut être exploité, pour étudier par exemple les problèmes de causalité et d'inversibilité d'un modèle d -varié à coefficients constants. Du fait que les coefficients (matrices) du modèle d -varié (2.5.4) sont donnés en terme des coefficients du modèle périodique (2.5.3) alors on obtient les conditions nécessaires et suffisantes que doivent satisfaire les paramètres du modèle périodique pour qu'il soit causal et inversible.

5.3. Condition de Causalité d'un modèle ARMA multivarié

Les conditions de causalité du modèle $\text{ARMA}(p^*, q^*)$ multivarié à coefficients constants,

$$\sum_{i=0}^p A_i X(t-i) = \sum_{j=0}^q B_j \varepsilon(t-j) \quad (2.5.8)$$

où $\varepsilon(t)$ est un processus multivarié non corrélé et les matrices A_i et B_j , sont des matrices carrées réelles. Sont données par Hannan (1970) et consistent à ce que les racines de l'équation :

$$\det \left[\sum_{i=0}^p A_i z^{p-i} \right] = 0; \quad z \in \mathbb{C} \quad (2.5.9)$$

sont à l'intérieur du cercle unité.

5.4. Condition d'invisibilité d'un modèle ARMA multivarié

Les conditions d'invisibilité du modèle multivarié à coefficients constants, $\text{ARMA}(p^*, q^*)$

$$\sum_{i=0}^{p^*} A_i X(t-i) = \sum_{j=0}^{q^*} B_j \varepsilon(t-j) \quad (2.5.10)$$

où $\varepsilon(t)$ est un processus multivarié non corrélé et les matrices A_i et B_j , sont des matrices carrées réelles; consistent à ce que les racines de l'équation :

$$\det \left[\sum_{i=0}^q B_j z^{q-j} \right] = 0; \quad z \in \mathbb{C} \quad (2.5.11)$$

sont à l'intérieur du cercle unité.

Remarque

Les conditions de causalité et d'inversibilité du modèle (2.5.3) découlent, directement, de celles du modèle multivarié qui lui est équivalent.

5.5. Condition d'inversibilité d'un modèle autorégressif périodique :

Le modèle univarié PAR(p_t) d -périodique :

$$y_t - \sum_{i=1}^{p_t} \phi_{ti} y_{t-i} = \varepsilon_t \quad t \in \mathbb{Z};$$

est causal si, et seulement si le modèle d -varié autorégressif à coefficients constants qui lui correspond :

$$\phi(0)X(\tau) + \phi(1)X(\tau-1) + \dots + \phi(p^*)X(\tau-p^*) = \eta(\tau) \quad (2.5.12)$$

est causal. C'est à dire, si et seulement si les racines de l'équation caractéristique

$$\det \left(\sum_{k=0}^{p^*} \Phi(k) z^{p^*-k} \right) = 0, \quad z \in \mathbb{C}$$

sont à l'intérieur du cercle unité, où les matrices $\Phi(k)$, $k = 1, \dots, p^*$ sont données par (2.5.6).

5.6. Condition d'inversibilité d'un modèle moyenne mobile périodique :

Le modèle univarié MA(q_t) d -périodique : $y_t = \varepsilon_t - \sum_{j=1}^{q_t} \theta_{tj} \varepsilon_{t-j}$; est inversible si et

seulement si le modèle d -varié moyenne mobile à coefficients constants qui lui correspond est inversible. C'est à dire, les racines de l'équation caractéristique

$$\det \left(\sum_{k=0}^{q^*} \Theta(k) z^{q^*-k} \right) = 0, \quad z \in \mathbb{C}$$

sont à l'intérieur du cercle unité, où les matrices $\Theta(k)$, $k = 1, \dots, q^*$ sont données par (2.5.7)

5.7. Technique « order-span lumping » et étude des modèles univariés autorégressifs moyenne mobile périodiques ARMA $_d(p_t, q_t)$

Dans cette section, nous présentons brièvement une autre approche, dite « order-span lumping », utilisé dans l'étude des propriétés théoriques des modèles de séries chronologiques périodiques (cf., Bentarzi (1998), Bentarzi et Hallin (1995) et (1998) et Ula et Smadi (1997)). Cette nouvelle approche consiste en la représentation d'un modèle univarié

(m -varié) moyenne mobile d -périodique d'ordre q par un modèle q (mq) varié moyenne mobile S -périodique d'ordre 1, où S est en fonction de d et q . ainsi le modèle obtenu est exploité pour étudier les propriétés d'un processus moyenne mobile périodique, en particulier les conditions d'inversibilité et le problème de la factorisation spectrale. Cette technique a été exploitée par Ula et Smadi (1997) pour établir une condition nécessaire et suffisante pour qu'un modèle autorégressif périodique soit causal.

5.8. Modèle multivarié moyenne mobile périodiques d'ordre 1 associé à un modèle multivarié moyenne mobile périodique d'ordre q :

Soit $\{y_t, t \in Z\}$ un processus m -varié moyenne mobile d'ordre q et d -périodique

$$y_t = \theta_{t,0}\varepsilon_t + \theta_{t,1}\varepsilon_{t-1} + \dots + \theta_{t,q}\varepsilon_{t-q} \quad t \in Z \quad (2.5.13)$$

où $\{\varepsilon_t\}$ est un processus bruit blanc de variance périodique et $\theta_{k,i}, i = 0, \dots, q$ sont des matrices ($m \times m$).

Bentarzi et Hallin (1995) considèrent le processus mq -varié

$$\begin{aligned} Y_T &= (y'_{qT}, y'_{qT-1}, \dots, y'_{q(T-1)+1})' \\ \eta_T &= (\varepsilon'_{qT}, \varepsilon'_{qT-1}, \dots, \varepsilon'_{q(T-1)+1})' \end{aligned}$$

Le modèle moyenne mobile périodique (2.5.13) se présente, en terme de processus $\{Y_T, T \in Z\}$ sous la forme suivante :

$$Y_T = \Theta_{T,0}\eta_T + \Theta_{T,1}\eta_{T-1} \quad (2.5.14)$$

Les matrices $\Theta_{T,0}$ et $\Theta_{T,1}$, de dimension ($mq \times mq$) sont données en fonction des coefficients périodiques $\theta_{t,i}, i = 0, \dots, q_t$ du modèle (2.5.13) comme suit :

$$\Theta_{T,0} = \begin{pmatrix} \theta_{qT,0} & \theta_{qT,1} & \dots & \theta_{qT,i-1} & \dots & \theta_{qT,q-1} \\ 0_m & \theta_{qT-1,0} & \theta_{qT-1,1} & \dots & \dots & \theta_{qT-1,q-2} \\ \vdots & \vdots & \dots & \ddots & \dots & \vdots \\ 0_m & 0_m & \dots & \theta_{qT-i+1,0} & \dots & \theta_{qT-i+1,q-i} \\ \vdots & \vdots & \dots & \dots & \ddots & \vdots \\ 0_m & & \dots & \dots & \dots & \theta_{qT-q+1,q} \end{pmatrix} \quad (2.5.15)$$

et

$$\Theta_{T,1} = \begin{pmatrix} \theta_{qT,q} & 0_m & \cdots & \cdots & \cdots & 0_m \\ \theta_{qT-1,q-1} & \theta_{qT-1,q} & 0_m & \cdots & \cdots & 0_m \\ \vdots & \vdots & \cdots & \ddots & \cdots & \vdots \\ \theta_{qT-i+1,q-i+1} & \cdots & \theta_{qT-i+1,q} & 0_m & \cdots & 0_m \\ \vdots & \vdots & \cdots & \cdots & \ddots & \vdots \\ \theta_{qT-q+1,1} & \cdots & \cdots & \cdots & \cdots & \theta_{qT-q+1,q} \end{pmatrix} \quad (2.5.16)$$

Le modèle (2.5.14) est un modèle mq -varié moyenne mobile S -périodique d'ordre 1, où S ($S > 1$) est le plus petit entier tel que : Sq est le plus petit multiple commun de q et d .

Proposition 2.5.1 (Bentarzi et Hallin (1995)) :

Le modèle m -varié moyenne mobile d -périodique d'ordre q , (2.5.13), est inversible si et seulement si les racines du polynôme :

$$|\Psi - Iz| = 0$$

sont à l'intérieur du cercle unité.

où Ψ est une matrice carrée, de dimension $(mq \times mq)$ donnée par :

$$\Psi = \Theta_{S,0}^{-1} \Theta_{S,1} \Theta_{S-1,0}^{-1} \Theta_{S-1,1} \cdots \Theta_{2,0}^{-1} \Theta_{2,1} \Theta_{1,0}^{-1} \Theta_{1,1}$$

où les matrices $\Theta_{j,0}$ et $\Theta_{j,1}$, $j = 1, \dots, S$ sont données respectivement par (2.5.15) et (2.5.16).

Nous remarquons que les éléments de la matrice Ψ sont exprimés en terme des coefficients d -périodiques du modèle périodique (2.5.13).

Dans le cas simple d'un modèle univarié moyenne mobile périodique d'ordre 1, la condition nécessaire et suffisante est donnée par le corollaire suivant :

Corollaire 2.5.1 (Bentarzi et Hallin(1995)) :

Le modèle univarié moyenne mobile d -périodique d'ordre 1 :

$$y_t = \theta_{t,0} \varepsilon_t + \theta_{t,1} \varepsilon_{t-1} \quad t \in \mathbb{Z}$$

est inversible si et seulement si :

$$\left| \frac{\theta_{1,1} \theta_{2,1} \cdots \theta_{d,1}}{\theta_{1,0} \theta_{2,0} \cdots \theta_{d,0}} \right| < 1.$$

Bentarzi (1998) a exploité cette nouvelle technique pour résoudre le problème de la factorisation spectrale d'un modèle m -varié moyenne mobile d -périodique.

Par analogie, Ula et Smadi (1997) ont exploité la technique introduite par Bentarzi et Hallin (1995) pour obtenir une condition simple et suffisante pour la stationnarité des modèles autorégressifs moyennes mobiles.

5.9. Modèle AR multivarié périodique d'ordre 1 associé à un modèle AR multivarié périodique d'ordre p :

Soit $\{y_t, t \in Z\}$ un processus périodiquement corrélé de période d satisfaisant au modèle m -varié autorégressif périodique de période d et d'ordre p et :

$$y_t = \phi_{t,1}y_{t-1} + \dots + \phi_{t,p}y_{t-p} + \varepsilon_t \quad (2.5.17)$$

Ula et Smadi (1997) définissent

$$\begin{aligned} Y_T &= (y'_{pT}, y'_{pT-1}, \dots, y'_{p(T-1)+1})' \\ \eta_T &= (\varepsilon'_{pT}, \varepsilon'_{pT-1}, \dots, \varepsilon'_{p(T-1)+1})' \end{aligned}$$

le modèle (2.5.17) s'écrira donc sous la forme suivante :

$$\Phi_{T,0}Y_T + \Phi_{T,1}Y_{T-1} = \eta_T \quad (2.5.18)$$

les matrices $\Phi_{T,0}$ et $\Phi_{T,1}$, de dimension $(mq \times mq)$ sont données en fonction des coefficients périodiques $\phi_{t,i}$, $i = 1, \dots, p$, $t \in Z$ du modèle (2.5.17) comme suit :

$$\Phi_{T,0} = \begin{pmatrix} I_m & -\phi_{Tp,1} & -\phi_{Tp,2} & \dots & -\phi_{Tp,p-1} \\ 0_m & I_m & -\phi_{Tp-1,1} & \dots & -\phi_{Tp-1,p-2} \\ \vdots & \dots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0_m & \dots & \dots & 0_m & I_m \end{pmatrix} \quad (2.5.19)$$

et

$$\Phi_{T,1} = \begin{pmatrix} -\phi_{Tp,p} & 0_m & 0_m & \dots & 0_m \\ -\phi_{Tp-1,p-1} & -\phi_{Tp-1,p} & 0_m & \dots & 0_m \\ \vdots & \dots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ -\phi_{Tp-p+1,1} & -\phi_{Tp-p+1,2} & \dots & \dots & -\phi_{Tp-p+1,p} \end{pmatrix} \quad (2.5.20)$$

Proposition 2.5.2 (Ula et Smadi (1997)) :

Le modèle m -varié autorégressif d -périodique d'ordre q (2.5.17) est inversible si et seulement si les racines du polynôme :

$$|\Omega - Iz| = 0$$

sont à l'intérieur du cercle unité, où Ω est la matrice carrée de dimension $(mq \times mq)$ donnée par :

$$\Omega = \Phi_{S,0}^{-1} \Phi_{S,1} \Phi_{S-1,0}^{-1} \Phi_{S-1,1} \cdots \Phi_{2,0}^{-1} \Phi_{2,1} \Phi_{1,0}^{-1} \Phi_{1,1}$$

où les matrices $\Phi_{j,0}$ et $\Phi_{j,1}$, $j = 1, \dots, S$ sont données par (2.5.19) et (2.5.20).

Ces deux derniers auteurs ont montré que l'utilisation de l'approche de Gladyshev (1961), pour déterminer les conditions de causalité, consiste à montrer que les valeurs propres d'une certaine matrice, de dimension $(mdp^* \times mdp^*)$, sont à l'intérieur du cercle unité. La nouvelle représentation (2.5.18) donne les mêmes conditions de causalité mais pour une matrice Ω de dimension $(mp \times mp)$, avec $mp \leq mdp^*$ (l'égalité est vérifiée lorsque $p = d$, $p = 2d \dots$).

Chapitre 3

Estimation hors-ligne d'un modèle autorégressif moyenne mobile Périodique

1. Introduction

L'objet du présent chapitre est de présenter les principaux résultats concernant l'estimation hors-ligne des paramètres d'un modèle autorégressif moyenne mobile d -périodique (PARMA). Nous présenterons une revue détaillée, mais non exhaustive, des différents algorithmes d'estimations disponibles. Comme il a été déjà signalé dans les chapitres précédents, les méthodes d'estimation hors-ligne peuvent être classées en deux groupes : les méthodes utilisant directement les données (méthode du maximum de vraisemblance, méthode des moindres carrés...) et les méthodes utilisant des transformations des données (méthode de Yule-Walker...). Nous commençons, en premier lieu, par étudier la classe de méthodes utilisant des transformations des données. Nous présenterons, à ce sujet deux méthodes particulières : la première est basée sur les équations de Yule-Walker périodiques et la seconde est la méthode récursive de Durbin-Levinson, ce type de méthodes est souvent appelé méthodes des moments. Notons que les estimations produites par cette classe de méthodes seront considérées comme bonnes valeurs initiales pour les méthodes itératives (méthodes utilisant directement les données) telles que la méthode du maximum de vraisemblance ou la méthode des moindres carrés. Nous passerons ainsi à l'exposé de cette

classe des méthodes qui utilisent directement les données. Les algorithmes de cette classe sont de nature itérative, elles nécessitent en effet, plusieurs évaluations de la fonction critère et une routine d'optimisation numérique. Le choix de l'algorithme d'optimisation est plus ou moins arbitraire puisque d'une part les différentes méthodes connues sont équivalentes et d'autre part le facteur prépondérant qui influe sur la rapidité des calculs et l'espace mémoire est bien le problème de l'évaluation de la fonction critère.

Nous avons partagé ce chapitre en cinq sections : Suite à cette succincte introduction, nous étudierons dans une deuxième section la classe des méthodes des moments, nous commençons ainsi par l'exposé des principaux résultats concernant l'estimation, par la méthode des moments, des paramètres d'un processus autorégressif pur d -périodique d'ordre p . Nous récapitulons, ensuite dans cette même section la méthode des moments pour l'estimation des paramètres d'un modèle PARMA général. Nous terminerons cette section par un exposé concis d'une méthode, récursive, plus efficace, pour l'estimation des paramètres d'un modèle autorégressif périodique pur. La troisième section sera consacrée à la présentation des différentes méthodes utilisant directement les données, à savoir la méthode des moindres carrés, la méthode du maximum de vraisemblance et leurs différentes variantes. Nous exposerons dans la section suivante un algorithme précis de Vecchia (1985a) pour l'estimation du maximum de vraisemblance d'un modèle PARMA. Nous terminons le chapitre par une application numérique en vérifiant les résultats obtenus à travers une étude de simulation intensive sur des séries artificielles supposées obtenues à partir de modèles bien spécifiés.

2. Méthodes basées sur des transformations des données dans les modèles ARMA périodiques (méthodes des moments)

L'estimation, par la méthode des moments, des paramètres d'un modèle PARMA est basée sur les équations de Yule-Walker généralisées au cas de modèles périodiques ou sur les équations récursives de prédiction. Le principe de la méthode des moments consiste à remplacer, dans les équations de Yule-Walker périodiques ou dans les équations récursives de prédiction (cf. chapitre 2), les autocovariances théoriques par leurs estimations empiriques. Celles-ci même sont obtenues grâce au remplacement des moments théoriques par leurs correspondants empiriques.

Supposons que nous disposons d'une série d'observations $\{\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_{Nd}\}$ considérée comme réalisation d'un modèle PARMA gaussien. Soit alors la série *standard* $\{y_1, y_2, \dots, y_{Nd}\}$ obtenue à partir de la première par correction de la moyenne empirique c'est à dire que : $y_{i+d\tau} = \tilde{y}_{i+d\tau} - \tilde{\mu}_i$, $\tau = 0, \dots, N-1$ et $i = 1, \dots, d$ où :

$$\tilde{\mu}_i = N^{-1} \sum_{\tau=0}^{N-1} \tilde{y}_{i+d\tau}, \quad i = 1, \dots, d \quad (3.2.1)$$

est un estimateur de la moyenne μ_i relative à la saison i .

Dans ce cas l'estimateur, par la méthode des moments, de la fonction d'autocovariance $\gamma_h^{(i)}$, pour la saison i et à l'horizon h , est la fonction d'autocovariance empirique notée $\tilde{\gamma}_h^{(i)}$, est donnée par (2.2.6) (cf. chapitre 2).

Théorème 3.2.1 (Convergence des estimateurs des moments)

Si y_1, y_2, \dots, y_{Nd} est une série d'observations générée par un modèle autorégressif d -périodique d'ordre p , alors les estimateurs $\tilde{\phi}$ et $\tilde{\sigma}$ définis par (3.2.4) sont convergent (quand $N \rightarrow +\infty$) presque sûrement. Et $N^{1/2}(\tilde{\phi}_k - \phi_k)$ $k = 1, \dots, d$ se distribue asymptotiquement selon une loi gaussienne de moyenne nulle et de matrice de covariance F^{-1} , où F est le bloc approprié de la matrice d'information donnée par le théorème 2.2.1 (cf. chapitre 2). Ainsi les estimateurs sont, dans ce sens, asymptotiquement efficaces.

où $\tilde{\phi}_k = (\tilde{\phi}_{k1}, \tilde{\phi}_{k2}, \dots, \tilde{\phi}_{kp})^T$.

Preuve voir (Pagano, 1978) \square

2.1.2. Cas de modèles autorégressifs moyennes mobiles périodiques

Lorsque nous rajoutons une composante moyenne mobile c'est à dire en considérant le modèle PARMA général, l'estimation par la méthode des moments devient plus délicate. En effet, pour les paramètres MA périodiques, nous aurons besoin de résoudre des systèmes non linéaires très compliqués. Pour le cas très particulier $q_i = 1$, $i = 1, \dots, d$ les équations obtenus par la méthode des moments ont été détaillées par (Salas et al., (1982)). Cependant pour notre présent cas, les estimateurs obtenus seront considérés seulement comme outils nécessaires (valeurs initiales) pour les méthodes itératives telles que du maximum de vraisemblance ou des moindres carrés. En effet cette classe de méthodes, comme nous verrons plus loin, requiert une routine d'optimisation numérique et par conséquent un ensemble d'itérations nécessaire jusque l'obtention de la convergence voulue. Donc il y a un grand intérêt d'utiliser des valeurs initiales convenables pour ne pas heurter la sensibilité de la méthode.

D'après Vecchia, (1985b), la valeur $\tilde{\phi}_i$ obtenue par résolution du système :

$$\tilde{\Gamma}_{i,q} \tilde{\phi}_i = \tilde{\gamma}_{i,q} \quad (3.2.5)$$

où : $(\tilde{\Gamma}_{i,q})_{kj} = \gamma_{j-k-q}^{(i-k-q)}$ ($k, j = 1, \dots, p$) et $\tilde{\gamma}_{i,q} = (\tilde{\gamma}_{q+1}^{(i)}, \tilde{\gamma}_{q+2}^{(i)}, \dots, \tilde{\gamma}_{q+p}^{(i)})^T$

Semble être une valeur initiale convenable.

La procédure du maximum de vraisemblance qui sera présentée dans la section 4 a montré une certaine relative insensibilité aux valeurs initiales des paramètres MA périodiques. Par conséquent, afin d'éviter la résolution des équations relatives aux paramètres MA périodiques qui, comme nous avons mentionné ci-dessus, engendre un coût de calcul considérable, ces paramètres prendront comme valeur initiale nécessaires aux estimations itératives la valeur zéro.

2.1.3. Exemple d'application

Pour illustrer l'estimation par la méthode des moments basée sur une série y_1, y_2, \dots, y_{Nd} . Considérons le modèle $AR_2(1)$ suivant :

$$\begin{cases} y_{1+2\tau} = \phi_1 y_{2\tau} + \varepsilon_{1+2\tau} \\ y_{2+2\tau} = \phi_2 y_{1+2\tau} + \varepsilon_{2+2\tau} \end{cases} \quad (3.2.6)$$

Soit $\gamma_h^{(i)} = \gamma(i, i-h)$ la covariance relative à la saison i pour l'horizon h . Multiplions (3.2.6) par $y_{i+2\tau-h}$, $i=1,2$, $h=0,1$; et prenons l'espérance mathématique des deux membres. Alors les équations de Yule-walker périodiques (ou saisonnières) sont données par :

$$\begin{cases} \gamma_0^{(1)} - \phi_1^2 \gamma_0^{(2)} = \sigma_1^2 \\ \gamma_0^{(2)} - \phi_2^2 \gamma_0^{(1)} = \sigma_2^2 \\ \gamma_1^{(1)} - \phi_1 \gamma_0^{(2)} = 0 \\ \gamma_1^{(2)} - \phi_2 \gamma_0^{(1)} = 0 \end{cases} \quad (3.2.7)$$

Les autocovariances empiriques $\tilde{\gamma}_h^{(i)}$, $i=1,2$; $h=0,1$ sont données par :

$$\begin{cases} \tilde{\gamma}_0^{(1)} = N^{-1} \sum_{\tau=0}^{N-1} y_{1+d\tau} y_{1+d\tau} \\ \tilde{\gamma}_0^{(2)} = N^{-1} \sum_{\tau=0}^{N-1} y_{2+d\tau} y_{2+d\tau} \\ \tilde{\gamma}_1^{(1)} = N^{-1} \sum_{\tau=0}^{N-1} y_{1+d\tau} y_{d\tau} \\ \tilde{\gamma}_1^{(2)} = N^{-1} \sum_{\tau=0}^{N-1} y_{2+d\tau} y_{1+d\tau} \end{cases} \quad (3.2.8)$$

En remplaçant dans (3.2.7) les $\gamma_h^{(i)}$ par les $\tilde{\gamma}_h^{(i)}$ et en résolvant le système obtenu par rapport à ϕ_1 , ϕ_2 , σ_1^2 , σ_2^2 , nous aurons les estimateurs des moments $\tilde{\phi}_1$, $\tilde{\phi}_2$, $\tilde{\sigma}_1^2$, $\tilde{\sigma}_2^2$, donnés par

$$\begin{pmatrix} \tilde{\phi}_1 \\ \tilde{\phi}_2 \\ \tilde{\sigma}_1^2 \\ \tilde{\sigma}_2^2 \end{pmatrix} = \begin{pmatrix} \tilde{\gamma}_1^{(1)} / \tilde{\gamma}_0^{(2)} \\ \tilde{\gamma}_1^{(2)} / \tilde{\gamma}_0^{(1)} \\ \tilde{\gamma}_0^{(1)} - \phi_1^2 \tilde{\gamma}_1^{(1)} \\ \tilde{\gamma}_0^{(2)} - \phi_2^2 \tilde{\gamma}_1^{(2)} \end{pmatrix} \quad (3.2.9)$$

2.2. Estimation des moments par des méthodes récursives

Au chapitre 2 nous avons vu que les coefficients du meilleur prédicteur linéaire de y_t basé sur les r variables aléatoires : $y_{t-r}, y_{t-r+1}, \dots, y_{t-1}$, où $\{y_t, t \in \mathbb{Z}\}$ est un processus périodiquement corrélé, s'obtiennent en résolvant un système d'équations de $d \times r$ variables. Nous verrons plus loin qu'il est possible d'obtenir ces coefficients par d'autres méthodes alternatives (méthodes récursives). Afin d'éviter le risque de toute confusion, notons qu'il existe deux types de méthodes récursives, le premier, que nous allons l'aborder dans la première partie de notre travail, concerne les méthodes non récursives vis-à-vis du temps mais récursives vis-à-vis de l'ordre du modèle. Le second type, qui sera l'objet des chapitres de la deuxième partie, concernera les méthodes récursives vis-à-vis du temps. L'objet de ce paragraphe est de présenter une méthode récursive (vis-à-vis de l'ordre du modèle) pour l'estimation des paramètres d'un modèle autorégressif périodique pur. Le principal profit tiré de cette méthode, spécialement pour les modèles d'ordres élevés, est la réduction du nombre d'opérations élémentaires (multiplications et divisions) qui sera de l'ordre de $O(d^2 \times p^2)$ au lieu de $O(d \times p^3)$ opérations pour la méthode directe (cf. Sakai, (1982)) après résolution des équations de Yule-Walker périodiques. A cause de la périodicité du processus $\{y_t, t \in \mathbb{Z}\}$ les

coefficients du meilleur prédicteur de y_t basé sur les p variables aléatoires : $y_{t-p}, y_{t-p+1}, \dots, y_{t-1}$, sont identiques aux paramètres (dans l'ordre convenable) d'un modèle autorégressif d'ordre p , d -périodique. Pour déterminer de manière récursive ces paramètres, l'idée principale est de déterminer progressivement les coefficients du meilleur prédicteur de y_t basé sur r variables aléatoires : $y_{t-r}, y_{t-r+1}, \dots, y_{t-1}$, et ce, pour r variant de 1 jusqu'à p .

Pour les équations récursives de ce type, deux ensembles de prédicteurs linéaires sont introduit. Le prédicteur linéaire prospectif et le prédicteur linéaire retrospectif. Le premier sera noté par un 'chapeaux', quant au second il sera noté par un 'tilde'. Le meilleur prédicteur linéaire (ou l'espérance conditionnelle linéaire) d'une variable y par rapport à l'ensemble des variables y_1, y_2, \dots est noté par $\hat{P}(y / y_1, y_2, \dots)$. Pour alléger cette écriture nous remplaçons l'ensemble des variables y_1, y_2, \dots par le sous-espace engendré par celle-ci, ce qui donne l'écriture suivante : $\hat{P}(y / M)$, où M est l'espace en question. Cette notation est similaire à celle utilisée pour l'espérance conditionnelle avec la lettre \hat{P} à la place de E . Ces quantités sont définies par :

$$\begin{aligned}\hat{y}_t^{(r)} &= \hat{P}(y_t / y_{t-1}, \dots, y_{t-r}) = \sum_{j=1}^r \phi_{ij}^{(r)} y_{t-j} \\ \sigma_t^{2(r)} &= \|y_t - \hat{y}_t^{(r)}\|^2 \\ \tilde{y}_t^{(r)} &= \hat{P}(y_t / y_{t+1}, \dots, y_{t+r}) = \sum_{j=1}^r \beta_{ij}^{(r)} y_{t+j} \\ v_t^{2(r)} &= \|y_t - \tilde{y}_t^{(r)}\|^2\end{aligned}$$

L'algorithme qui suit est la généralisation de l'algorithme de Durbin-Levinson (Durbin (1960)), aux cas des modèles autorégressifs d -périodiques, par Boshnakov (1996). Notons que des équations équivalentes construites par une autre approche se trouvent dans Sakai, (1982).

Proposition 3.2.1

Considérons le modèle autorégressif périodique suivant :

$$Z_{i+d\tau} - \sum_{j=1}^p \phi_{ij} Z_{i+d\tau-j} = \varepsilon_{i+d\tau} \quad , i = 1, \dots, d ; \tau \in \mathbb{Z}. \quad (3.2.10)$$

Notons par ϕ_i $i = 1, \dots, d$ le $P \times 1$ vecteur des paramètres pour la saison i défini par $\phi_i = (\phi_{i1}, \phi_{i2}, \dots, \phi_{iP})^T$. Et supposons que le processus est causal. Alors le schéma récursif pour le calcul de ϕ_{ij} à partir de $\gamma_h^{(i)}$, $i = 1, \dots, d$, $j, h = 1, \dots, P$ est donné par Boshnakov (1994a, b) et contenant les équations suivantes :

$$\phi_{ir}^{(r)} = \left(\gamma_r^{(i)} - \sum_{j=1}^{r-1} \beta_{i-r,j}^{(r-1)} \gamma_{r-j}^{(i)} \right) / v_{i-r}^{2(r-1)} \quad (3.2.11)$$

$$\phi_{ij}^{(r)} = \phi_{ij}^{(r-1)} - \phi_{ir}^{(r)} \beta_{i-r,r-j}^{(r-1)} \quad j = 1, \dots, r-1 \quad (r \geq 1) \quad (3.2.12)$$

$$\sigma_i^{2(r)} = \sigma_i^{2(r-1)} - (\phi_{ir}^{(r)})^2 v_{i-r}^{2(r-1)} \quad (3.2.13)$$

$$\beta_{ir}^{(r)} = \left(\gamma_{-r}^{(i)} - \sum_{j=1}^{r-1} \phi_{i+r,j}^{(r-1)} \gamma_{j-r}^{(i)} \right) / \sigma_{i+r}^{2(r-1)} \quad (3.2.14)$$

$$\beta_{ij}^{(r)} = \beta_{ij}^{(r-1)} - \beta_{ir}^{(r)} \phi_{i+r,r-j}^{(r-1)} \quad j = 1, \dots, r-1 \quad (r \geq 1) \quad (3.2.15)$$

$$v_i^{2(r)} = v_i^{2(r-1)} - (\beta_{ir}^{(r)})^2 \sigma_{i+r}^{2(r-1)} \quad (3.2.16)$$

Ces six équations seront appliquées dans l'ordre donné pour $k = 1, \dots, P$, $i = 1, \dots, d$, en commençant par $\sigma_i^{2(0)} = v_i^{2(0)} = \gamma_0^{(i)}$, $i = 1, \dots, d$, la première étape donne :

$$\phi_{i1}^{(1)} = \gamma_1^{(i)} / \gamma_0^{(i)} \quad \text{et} \quad \beta_{i1}^{(1)} = \gamma_{-1}^{(i)} / \gamma_0^{(i)}. \quad \text{Par la suite les coefficients autorégressifs seront obtenus à partir de}$$

$$\phi_{ij} = \phi_{ij}^{(P)} \quad \text{et} \quad \sigma_i^2 = \sigma_i^{2(P)}, \quad i = 1, \dots, d. \quad (3.2.17)$$

Cette procédure est fréquemment utilisée lorsqu'une estimation des autocovariances est disponible. Si les $\gamma_h^{(i)}$ sont les autocovariances d'un processus périodiquement corrélé alors $\phi_{ij}^{(r)}$, $i = 1, \dots, d$, $j = 1, \dots, r$ sont les coefficients du meilleur prédicteur linéaire de y_t basé sur les r variables y_{t-1}, \dots, y_{t-r} , et $\sigma_t^{2(r)}$ est la variance de l'erreur de prédiction. Les $\beta_{ij}^{(r)}$ $i = 1, \dots, d$, $j = 1, \dots, r$ sont, quant à eux, les coefficients du meilleur prédicteur linéaire de y_t basé sur y_{t+1}, \dots, y_{t+r} , et $v_t^{2(r)}$ est la variance de l'erreur de prédiction correspondante.

Preuve

La démonstration de cette proposition n'a pas été donnée dans Boshnakov (1996) cependant, en s'inspirant de la démonstration de Brockwel et Davis (1986) pour le cas classique nous avons établi cette preuve.

Notons par : $M_{t,r}, t \in Z, r \in N - \{0\}$, le sous-espace de Hilbert engendré par les r variables aléatoires $y_{t-r}, y_{t-r+1}, \dots, y_{t-1}$, $M_{t,r-1}, t \in Z$, le sous-espace engendré par les $r-1$ variables $y_{t-r+1}, y_{t-r+2}, \dots, y_{t-1}$ et $M_{t,r-1}^\perp, t \in Z$, le sous-espace engendré par la variable $y_{t-r} - \hat{P}(y_{t-r} / M_{t,r-1})$ où $\hat{P}(y / A)$ note la projection orthogonale du vecteur y dans le sous espace A . Comme les sous-espaces $M_{t,r-1}$ et $M_{t,r-1}^\perp$ sont orthogonaux dans $M_{t,r}$ alors nous pouvons écrire

$$\hat{P}(y_t / M_{t,r}) = \hat{P}(y_t / M_{t,r-1}) + \hat{P}(y_t / M_{t,r-1}^\perp) \quad (3.2.18)$$

or, nous avons

$$\hat{y}_t^{(r)} = \hat{P}(y_t / M_{t,r}) = \sum_{j=1}^r \phi_{ij}^{(r)} y_{t-j}$$

$$\hat{P}(y_t / M_{t,r-1}) = \sum_{j=1}^{r-1} \phi_{ij}^{(r-1)} y_{t-j}$$

$$\hat{P}(y_{t-r} / M_{t,r-1}) = \sum_{j=1}^{r-1} \beta_{t-r,j}^{(r-1)} y_{t-r+j}$$

$$\hat{P}(y_t / M_{t,r-1}^\perp) = a(y_{t-r} - \hat{P}(y_{t-r} / M_{t,r-1}))$$

avec a est une constante qui s'obtienne au moyen de l'équation suivante :

$$a = \frac{\langle y_t, y_{t-r} - \hat{P}(y_{t-r} / M_{t,r-1}) \rangle}{\|y_{t-r} - \hat{P}(y_{t-r} / M_{t,r-1})\|^2} \quad (3.2.19)$$

(3.2.18) devient

$$\hat{P}(y_t / M_{t,r}) = \hat{P}(y_t / M_{t,r-1}) + a(y_{t-r} - \hat{P}(y_{t-r} / M_{t,r-1}))$$

par la suite nous aurons

$$\sum_{j=1}^r \phi_{ij}^{(r)} y_{t-j} = \sum_{j=1}^{r-1} \phi_{ij}^{(r-1)} y_{t-j} + a \left(y_{t-r} - \sum_{j=1}^{r-1} \beta_{t-r,j}^{(r-1)} y_{t-r+j} \right)$$

ou encore

$$ay_{t-r} + \sum_{j=1}^{r-1} (\phi_{ij}^{(r-1)} - a\beta_{t-r,j}^{(r-1)}) y_{t-j} = \phi_{ij}^{(r)} + \sum_{j=1}^{r-1} \phi_{ij}^{(r)} y_{t-j}$$

Par identification il s'en suit que

$$\begin{cases} \phi_{tr}^{(r)} = a \\ \phi_{ij}^{(r)} = \phi_{ij}^{(r-1)} - a\beta_{t-r,r-j}^{(r-1)}, \quad j = 1, \dots, r-1 \end{cases}$$

En explicitant (3.2.19), a peut se mettre comme suit

$$a = \langle y_t, y_{t-r} \rangle - \left\langle y_t, \sum_{j=1}^{r-1} \beta_{t-r,j}^{(r-1)} y_{t-r+j} \right\rangle / v_{t-r}^{2(r-1)}$$

d'où

$$a = \left(\gamma_r^{(t)} - \sum_{j=1}^{r-1} \beta_{t-r,j}^{(r-1)} \gamma_{r-j}^{(t)} \right) / v_{t-r}^{2(r-1)}$$

ii) Par définition,

$$\begin{aligned} \sigma_t^{2(r)} &= \|y_t - \hat{y}_t^{(r)}\|^2 = \|y_t - \hat{P}(y_t / M_{t,r})\|^2 \\ &= \|y_t - \hat{P}(y_t / M_{t,r})\|^2 = \|y_t - \hat{P}(y_t / M_{t,r-1}) - \hat{P}(y_t / M_{t,r-1}^\perp)\|^2 \\ &= \|y_t - \hat{P}(y_t / M_{t,r-1})\|^2 + \|\hat{P}(y_t / M_{t,r-1}^\perp)\|^2 - 2\langle y_t - \hat{P}(y_t / M_{t,r-1}), \hat{P}(y_t / M_{t,r-1}^\perp) \rangle \\ &= \sigma_t^{2(r-1)} + a^2 v_{t+r}^{2(r-1)} - 2a \langle y_t, y_{t-r} - \hat{P}(y_{t-r} / M_{t,r-1}) \rangle \end{aligned}$$

or d'après (3.2.19) nous avons $\langle y_t, y_{t-r} - \hat{P}(y_{t-r} / M_{t,r-1}) \rangle = av_{t-r}^{2(r-1)}$

d'où

$$\sigma_t^{2(r)} = \sigma_t^{2(r-1)} - a^2 v_{t+r}^{2(r-1)} = \sigma_t^{2(r-1)} - (\phi_{tr}^{(r)})^2 v_{t+r}^{2(r-1)}.$$

iii) Notons maintenant par : $M_{t+r,r}, t \in \mathbb{Z}, r \in \mathbb{N} - \{0\}$, le sous-espace de Hilbert engendré par les r variables aléatoires $y_{t+1}, y_{t+2}, \dots, y_{t+r}$, $M_{t+r,r-1}, t \in \mathbb{Z}$, le sous-espace engendré par les $r-1$ variables $y_{t+1}, y_{t+2}, \dots, y_{t+r-1}$ et $M_{t+r,r-1}^\perp, t \in \mathbb{Z}$, le sous-espace engendré par la variable $y_{t+r} - \hat{P}(y_{t+r} / M_{t+r,r-1})$. Comme les sous-espaces $M_{t+r,r-1}$ et $M_{t+r,r-1}^\perp$ sont orthogonaux dans $M_{t+r,r}$ alors nous pouvons donc écrire

$$\begin{aligned} \tilde{y}_t^{(r)} &= \hat{P}(y_t / M_{t+r,r}) = \hat{P}(y_t / M_{t+r,r-1}) + \hat{P}(y_t / M_{t+r,r}^\perp) \\ &= \sum_{j=1}^{r-1} \beta_{t+r,j}^{(r-1)} y_{t+j} + b \left(y_{t+r} - \sum_{j=1}^{r-1} \phi_{t+r,j}^{(r-1)} y_{t+r-j} \right) \\ &= by_{t+r} + \sum_{j=1}^{r-1} (\beta_{t+r,j}^{(r-1)} - b\phi_{t+r,r-j}^{(r-1)}) y_{t+r-j} \end{aligned} \quad (3.2.20)$$

Ainsi nous obtenons l'identité suivante :

$$\beta_{tr}^{(r)} y_{t+r} + \sum_{j=1}^{r-1} \beta_{t+r,j}^{(r)} y_{t+j} = by_{t+r} + \sum_{j=1}^{r-1} (\beta_{t+r,j}^{(r-1)} - b\phi_{t+r,r-j}^{(r-1)}) y_{t+r-j}$$

dont la solution est donnée par

$$\begin{cases} \beta_{ir}^{(r)} = b \\ \beta_{ij}^{(r)} = \beta_{ij}^{(r-1)} - b\phi_{i+r,r-j}^{(r-1)}, \quad j = 1, \dots, r-1 \end{cases}$$

où b est une constante qui s'obtient au moyen de l'équation suivante :

$$b = \frac{\langle y_t, y_{t+r} - \hat{P}(y_{t+r} / M_{t+r,r-1}) \rangle}{\|y_{t+r} - \hat{P}(y_{t+r} / M_{t+r,r-1})\|^2} \quad (3.2.21)$$

$$b = \left(\langle y_t, y_{t+r} \rangle - \left\langle y_t, \sum_{j=1}^{r-1} \phi_{i+r,j}^{(r-1)} y_{t+r-j} \right\rangle \right) / \sigma_{t+r}^{2(r-1)}$$

d'où

$$\beta_{ir}^{(r)} = b = \left(\gamma_{-r}^{(t)} - \sum_{j=1}^{r-1} \phi_{i+r,j}^{(r-1)} \gamma_{j-r}^{(t)} \right) / \sigma_{t+r}^{2(r-1)}$$

iv) Nous avons déjà que

$$\begin{aligned} v_t^{2(r)} &= \|y_t - \tilde{y}_t^{(r)}\|^2 = \|y_t - \hat{P}(y_t / M_{t+r,r})\|^2 \\ &= \|y_t - \hat{P}(y_t / M_{t+r,r-1}) - \hat{P}(y_t / M_{t+r,r-1}^\perp)\|^2 \\ &= \|y_t - \hat{P}(y_t / M_{t+r,r-1})\|^2 + \|\hat{P}(y_t / M_{t+r,r-1}^\perp)\|^2 - 2\langle y_t - \hat{P}(y_t / M_{t+r,r-1}), \hat{P}(y_t / M_{t+r,r-1}^\perp) \rangle \\ &= v_t^{2(r-1)} + b^2 \sigma_{t+r}^{2(r-1)} - 2b \langle y_t, y_{t+r} - \hat{P}(y_{t+r} / M_{t+r,r-1}) \rangle \end{aligned}$$

or d'après (3.2.21) nous avons $\langle y_t, y_{t+r} - \hat{P}(y_{t+r} / M_{t+r,r-1}) \rangle = b \sigma_{t+r}^{2(r-1)}$

d'où

$$v_t^{2(r)} = v_t^{2(r-1)} - b^2 \sigma_{t+r}^{2(r-1)} = v_t^{2(r-1)} - (\beta_{ir}^{(r)})^2 \sigma_{t+r}^{2(r-1)}.$$

v) Comme les coefficients sont périodiques en t donc pour $t = i + d\tau$ nous pouvons remplacer dans les équations précédentes t par i . i variera de 1 jusqu'à d .

vi) Notons aussi que les équations récursives précédentes établissent le lien existant entre les coefficients du meilleur prédicteur d'une certaine variable y_t basé sur les r variables précédentes du processus y_{t-1}, \dots, y_{t-r} , et ceux du meilleur prédicteur linéaire de la même variable basé sur les $r-1$ précédentes $y_{t-1}, \dots, y_{t-r+1}$. Donc pour un modèle AR_d d'ordre p les paramètres ϕ_{ij} , $i = 1, \dots, d$, $j = 1, \dots, p$ seront les coefficients du meilleur prédicteur linéaire d'une variable $y_{i+d\tau}$ basé sur les p variables précédentes. Ainsi pour obtenir les estimations nous écrivons les équations récursives de prédiction pour $r = 1, \dots, p$. Ce qui achèvera la démonstration.

Remarque Les équations (3.2.13) et (3.2.16) peuvent être remplacés de manière équivalente par les équations suivantes (Boshnakov, (1996))

$$\sigma_i^{2(r)} = (1 - \phi_{ir}^{(r)} \beta_{i-r,r}^{(r)}) \sigma_i^{2(r-1)} \quad (3.2.13a)$$

$$v_i^{2(r)} = (1 - \phi_{i+r,r}^{(r)} \beta_{i,r}^{(r)}) v_i^{2(r-1)} \quad (3.2.16a)$$

En effet en utilisant la formule de Burg (Burg, (1975)) appliquée aux modèles autorégressifs périodiques par Sakai, (1982) nous retrouvons directement le résultat.

3. Estimation dans les modèles PARMA par des méthodes utilisant directement les données

Comme pour le cas des modèles ARMA classiques, ce qui a déjà été présenté au chapitre 1, nous traiterons dans cette section le problème de l'estimation des paramètres d'un modèle ARMA périodique selon deux grands critères, la méthode des moindres carrés et la méthode du maximum de vraisemblance. Nous commençons tout d'abord par dénombrer les différents critères pris en compte.

3.1. Critères

Le but est d'estimer les paramètres d'un modèle $PARMA_d(p,q)$ à partir d'une réalisation de longueur Nd . Nous avons envisagé les quatre critères suivants :

- Le critère des moindres carrés conditionnels (MCC).
- Le critère des moindres carrés non conditionnel (MCN).
- Le critère du maximum de vraisemblance conditionnel (MVC).
- Le critère du maximum de vraisemblance non conditionnel (MV).

Nous verrons par la suite que les critères (MCC) et (MVC) peuvent se confondre. Nous illustrerons essentiellement sur le processus $PARMA_d(p,q)$ suivant :

$$y_{i+d\tau} - \phi_{i1}y_{i+d\tau-1} - \phi_{i2}y_{i+d\tau-2} - \dots - \phi_{ip}y_{i+d\tau-p} = \varepsilon_{i+d\tau} - \theta_{i1}\varepsilon_{i+d\tau-1} - \dots - \theta_{iq}\varepsilon_{i+d\tau-q} \quad i=1,2,\dots,d, \quad (3.3.1)$$

Supposons que nous désirons estimer les paramètres de ce modèle en se basant sur la réalisation donnée. Alors ces estimateurs seront choisis de façon à minimiser (ou maximiser) la fonction critère fixée. Comme généralement cette fonction est non linéaire donc nous devons utiliser des routines d'optimisations itératives pour la minimiser, par conséquent plusieurs évaluations seront nécessaires. Il faut d'abord se pencher sur l'évaluation de la fonction critère. Le problème d'optimisation a déjà été abordé dans le chapitre 1.

3.2. Evaluation des fonctions critères

Pour une réalisation y_1, y_2, \dots, y_{Nd} le modèle (3.3.1) peut se mettre sous la forme suivante :

$$Ay_N = B\varepsilon_N + CI_* \quad (3.3.2)$$

où $y_N = (y_1, y_2, \dots, y_N)^T$, $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{Nd})^T$ et $I_* = (y_0, \dots, y_{1-p}, \varepsilon_0, \dots, \varepsilon_{1-q})^T$.

Notons que A et B sont des matrices triangulaires inférieures de taille $(Nd \times Nd)$ à diagonale unité. La matrice C , quant à elle, est de taille $(Nd \times (p+q))$ où seulement les $h (= \max\{p, q\})$ premières lignes sont non nulles. L'expression exacte de ces matrices est donnée par (voir Vecchia (1985a)) pour le cas de modèles moyenne mobile périodique) :

$$A = \begin{pmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ -\phi_{21} & 1 & 0 & \dots & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ -\phi_{p,p-1} & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & -\phi_{p,p-1} & \dots & -\phi_{p,1} & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ -\theta_{21} & 1 & 0 & \dots & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ -\theta_{q,q-1} & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & -\theta_{q,q-1} & \dots & -\theta_{q1} & 1 \end{pmatrix}$$

$$C_{(N \times (p+q))} = \begin{pmatrix} -\phi_{1p} & \cdots & \cdots & -\phi_{11} & \theta_{1,q} & \cdots & \cdots & \theta_{11} \\ 0 & -\phi_{2p} & \cdots & -\phi_{22} & 0 & \theta_{2,q} & \cdots & \theta_{22} \\ \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\phi_{pp} & \vdots & \ddots & \ddots & \theta_{q,q} \\ \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 0 & \cdots & \cdots & 0 \end{pmatrix} \quad (3.3.3)$$

Ces matrices peuvent s'écrire sous la forme condensée suivante

$$A_{ij} = \begin{cases} -[\phi_{i,i-j}] & \text{si } j < i \\ 1 & \text{si } j = i \\ 0 & \text{si } j > i \end{cases} \quad i, j = 1, \dots, Nd$$

$$B_{ij} = \begin{cases} -[\theta_{i,i-j}] & \text{si } j < i \\ 1 & \text{si } j = i \\ 0 & \text{si } j > i \end{cases} \quad i, j = 1, \dots, Nd$$

$$C_{ij} = \begin{cases} [\phi_{i,p+i-j}] & \text{si } j \leq i \text{ et } j \leq p \\ -[\theta_{i,p+q+i-j}] & \text{si } i \leq j - p \text{ et } j > p \\ 0 & \text{ailleurs} \end{cases} \quad i = 1, \dots, Nd, \quad j = 1, \dots, p + q$$

où

$$[\phi_{ij}] = \begin{cases} \phi_{ij} & \text{si } j = 1, \dots, p \\ 0 & \text{ailleurs} \end{cases}, \quad [\theta_{ij}] = \begin{cases} \theta_{ij} & \text{si } j = 1, \dots, q \\ 0 & \text{ailleurs} \end{cases}$$

Pour plus de clarté considérons la série y_1, y_2, \dots, y_{Nd} supposée issue du modèle MA₂(1) suivant :

$$\begin{cases} y_{1+2\tau} = \varepsilon_{1+2\tau} - \theta_1 \varepsilon_{2\tau} \\ y_{2+2\tau} = \varepsilon_{2+2\tau} - \theta_2 \varepsilon_{1+2\tau} \end{cases} \quad (3.3.4)$$

où $\theta = (\theta_1, \theta_2)^T$ est le vecteur des paramètres à estimer. Sur la base de l'observation donnée nous allons exprimer explicitement l'erreur ε_t à l'instant t en fonction des paramètres et des observations et ce pour chaque critère.

3.2.1. Somme des carrés conditionnelle

Ce critère consiste à choisir un estimateur des paramètres du modèle minimisant la somme des carrés. Contrairement au cas ARMA classique les membres de la somme seront pondérés par la variance de l'erreur. Nous supposons que $I_* = \varepsilon_0$ est fixé, soit par exemple égal à 0. Nous pouvons donc déterminer à partir de (3.3.4) $\varepsilon_1 = y_1$. puis

$$\varepsilon_2 = y_2 + \theta_2 y_1$$

$$\varepsilon_3 = y_3 + \theta_1 y_2 + \theta_2 \theta_2 y_1$$

et ainsi de suite l'erreur ε_t s'exprime en fonction de la parité de t . Pour $\tau \geq 0$ nous trouverons

$$\varepsilon_{1+d\tau} = \sum_{j=0}^{2\tau} \theta_1^{[(j+1)/2]} \theta_2^{[j/2]} y_{1+2\tau-j} \quad (3.3.5a)$$

$$\varepsilon_{2+d\tau} = \sum_{j=0}^{1+2\tau} \theta_1^{[j/2]} \theta_2^{[(j+1)/2]} y_{2+2\tau-j} \quad (3.3.5b)$$

où $[x]$ désigne la partie entière de x .

Nous avons pu expliciter la valeur de ε_t en fonction de $(\theta_1, \theta_2, \sigma_1^2, \sigma_2^2)$ et de y seulement. Par suite la somme des carrés (S_C) conditionnelle à $I_* = 0$ s'écrit comme suit :

$$\begin{aligned} S_C(\theta_1, \theta_2, \sigma_1^2, \sigma_2^2) &= \sum_{t=1}^{Nd} \frac{\varepsilon_t^2}{\sigma_t^2} = \sum_{i=1}^2 \sum_{\tau=0}^{N-1} \frac{\varepsilon_{i+2\tau}^2(\theta_1, \theta_2, \sigma_1^2, \sigma_2^2)}{\sigma_{i+2\tau}^2} \\ &= \sum_{\tau=0}^{N-1} \frac{\varepsilon_{1+2\tau}^2(\theta_1, \theta_2, \sigma_1^2, \sigma_2^2)}{\sigma_1^2} + \sum_{\tau=0}^{N-1} \frac{\varepsilon_{2+2\tau}^2(\theta_1, \theta_2, \sigma_1^2, \sigma_2^2)}{\sigma_2^2} \\ &= \sum_{\tau=0}^{N-1} \frac{\left(\sum_{j=0}^{2\tau} \theta_1^{[(j+1)/2]} \theta_2^{[j/2]} y_{1+2\tau-j} \right)^2}{\sigma_1^2} + \sum_{\tau=0}^{N-1} \frac{\left(\sum_{j=0}^{1+2\tau} \theta_1^{[j/2]} \theta_2^{[(j+1)/2]} y_{2+2\tau-j} \right)^2}{\sigma_2^2} \end{aligned} \quad (3.3.6)$$

Nous pouvons remarquer que la fonction (S_C) n'est pas une fonction quadratique en $(\theta_1, \theta_2, \sigma_1^2, \sigma_2^2)$. L'équation normale permettant d'estimer $(\theta_1, \theta_2, \sigma_1^2, \sigma_2^2)$ n'est donc pas linéaire.

Notons qu'en pratique nous ne pouvons pas toujours évaluer la fonction (S_C) explicitement en fonction des paramètres et des observations comme dans (3.3.5). Mais la valeur de l'erreur ε_t sera déduite à partir du schéma récursif.

3.2.2 Somme des carrés non conditionnelle

Ce critère se diffère du premier par le fait que nous devons déterminer une valeur de $I_* = \varepsilon_0$ meilleure que 0. Si le modèle (3.3.1) est correct, nous pouvons écrire aussi

$$\begin{cases} y_{1+2\tau} = \varepsilon'_{1+2\tau} - \theta_1 \varepsilon'_{2+2\tau} \\ y_{2+2\tau} = \varepsilon'_{2+2\tau} - \theta_2 \varepsilon'_{3+2\tau} \end{cases} \quad (3.3.7)$$

où les $\{\varepsilon'_t\}$ sont les innovations en temps inversé. En supposant $\varepsilon'_{2N+k} = 0$, pour $k \geq 1$, nous déterminons successivement $\varepsilon'_{2N} = y_{2N}$ puis $\varepsilon'_{2N-1} = y_{2N-1} + \theta_1 y_{2N}$

De façon similaire au (3.3.6) nous obtenons :

$$\varepsilon'_{2N-2} = y_{2N-2} + \theta_2 y_{2N-1} + \theta_1 \theta_2 y_{2N}$$

⋮

$$\varepsilon'_1 = \sum_{j=1}^{2N} \theta_1^{[j/2]} \theta_2^{[(j-1)/2]} y_j$$

ce qui fournit finalement la prévision rétrospective $\hat{y}_0 = -\theta_2 \varepsilon'_1$ qui se met sous la relation suivante

$$\hat{y}_0 = - \sum_{j=1}^{2N} \theta_1^{[j/2]} \theta_2^{[(j+1)/2]} y_j$$

En prenant maintenant $\varepsilon_0 = \hat{y}_0$ nous pouvons obtenir les ε_t par (3.3.6) seulement en fonction de $(\theta_1, \theta_2, \sigma_1^2, \sigma_2^2)$, ε_0 et de y . Nous en déduisons alors la somme des carrés non conditionnelle, soit

$$\begin{aligned} S_N(\theta_1, \theta_2, \sigma_1^2, \sigma_2^2) &= \sum_{t=0}^N \frac{\varepsilon_t^2}{\sigma_t^2} \\ &= \sum_{\tau=0}^{N-1} \sum_{j=0}^{2\tau} \frac{\theta_1^{[(j+1)/2]} \theta_2^{[j/2]} y_{1+2\tau-j}}{\sigma_1^2} + \sum_{\tau=0}^{N-1} \sum_{j=0}^{1+2\tau} \frac{\theta_1^{[j/2]} \theta_2^{[(j+1)/2]} y_{2+2\tau-j}}{\sigma_2^2} - \sum_{j=1}^{2N} \theta_1^{[j/2]} \theta_2^{[(j+1)/2]} y_j \end{aligned} \quad (3.3.8)$$

3.2.3 Fonction de vraisemblance conditionnelle

Comme nous allons voir plus loin, ce critère se confond avec celui du critère MCC. Nous avons donc besoin d'une hypothèse très précise sur la distribution liée des $\{\varepsilon_t\}$. Ce sont des variables aléatoires indépendants, de loi $N(0, \sigma_t^2)$. Considérons d'abord la fonction de vraisemblance $l_C(\theta_1, \theta_2)$ conditionnelle à $\varepsilon_0 = 0$. La fonction densité du vecteur $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{2N})^T$ est

$$f(\varepsilon) = \left(1/\sqrt{2\pi}\right)^{2N} \left(\prod_{t=1}^{2N} \sigma_t^2\right)^{-1/2} \exp\left[-\sum_{t=1}^{2N} \frac{\varepsilon_t^2}{2\sigma_t^2}\right].$$

ou encore

$$f(\varepsilon) = \left(1/\sqrt{2\pi}\right)^{2N} \sigma_1^{-N} \sigma_2^{-N} \exp\left\{-\sum_{\tau=0}^{N-1} \frac{\varepsilon_{1+2\tau}^2}{2\sigma_1^2} + \sum_{\tau=0}^{N-1} \frac{\varepsilon_{2+2\tau}^2}{2\sigma_2^2}\right\}. \quad (3.3.9)$$

Notons par $y = (y_1, y_2, \dots, y_{2N})^T$ le vecteur des observations. La matrice Jacobienne de la transformation de R^{2N} qui applique y sur ε à l'aide de l'équation (3.3.2) est triangulaire inférieure et les éléments diagonaux sont égaux à 1. La fonction de densité de y conditionnellement à $\varepsilon_0 = 0$ est donc donnée par (3.3.9) où ε est exprimé en fonction de y au moyen des relations (3.3.2). Nous aurons donc

$$\log L_C(\theta_1, \theta_2) = -N/2 \log(2\pi) - N/2 \log \sigma_1^2 - N/2 \log \sigma_2^2 - \sum_{\tau=0}^{N-1} \frac{\varepsilon_{1+2\tau}^2}{2\sigma_1^2} - \sum_{\tau=0}^{N-1} \frac{\varepsilon_{2+2\tau}^2}{2\sigma_2^2} \quad (3.3.10)$$

L'équation de vraisemblance correspondant à σ_1^2 et σ_2^2 fournit les deux équations suivantes :

$$\begin{cases} \sigma_1^2 = 1/N \sum_{\tau=0}^{N-1} \varepsilon_{1+2\tau}^2 \\ \sigma_2^2 = 1/N \sum_{\tau=0}^{N-1} \varepsilon_{2+2\tau}^2 \end{cases} \quad (3.3.11)$$

par substitution dans (3.3.10), nous obtenons donc à une constante additive près

$$-N/2 \log \sum_{\tau=0}^{N-1} \varepsilon_{1+2\tau}^2 - N/2 \log \sum_{\tau=0}^{N-1} \varepsilon_{2+2\tau}^2 \quad (3.3.12)$$

Ce qui montre que l'estimateur du maximum de vraisemblance $(\hat{\theta}_1, \hat{\theta}_2)$ de (θ_1, θ_2) est tel que l'expression de (3.3.12) est minimum. Connaissant $(\hat{\theta}_1, \hat{\theta}_2)$ nous déterminons l'estimateur $\hat{\sigma}_i^2$ de σ_i^2 par (3.3.11).

Remarque

Pour un modèle $\text{PARMA}_d(p, q)$ générale l'équation de vraisemblance pour σ_i^2 $i = 1, \dots, d$, fournit :

$$\sigma_i^2 = 1/N \sum_{\tau=0}^{N-1} \varepsilon_{i+d\tau}^2 \quad i = 1, \dots, d,$$

ainsi, par substitution dans l'équation générale, l'estimateur du maximum de vraisemblance conditionnel est tel que la fonction $\log\left(\prod_{i=1}^d \sum_{\tau=0}^{N-1} \varepsilon_{i+d\tau}^2\right)$ est minimum

3.2.4 Fonction de vraisemblance non conditionnelle

Comme pour le cas ARMA classique, le critère du MVC ne coïncide pas avec le critère MCN du 3.2.1 contrairement au cas conditionnel.

D'après (3.3.2) il vient que

$$y = A^{-1}B\varepsilon + A^{-1}CI_*$$

$$\text{cov}(y) = A^{-1}B \text{cov}(\varepsilon)B^T A^{-1T} + A^{-1}C \text{cov}(I_*)C^T A^{-1T}$$

or nous avons

$$\text{cov}(\varepsilon) = \text{Diag}(\sigma_1^2, \dots, \sigma_{Nd}^2)I_{Nd} = D_{1,Nd}^2 I_{Nd}.$$

(I_{Nd} est la matrice identité d'ordre Nd , $D_{i,j}^k$ est la matrice diagonale d'éléments $\sigma_i^k, \dots, \sigma_j^k$).

Alors pour des raisons de commodité notons par Γ la matrice de covariance du vecteur $\text{Diag}(\sigma_1^{-1}, \dots, \sigma_{Nd}^{-1})y$ alors nous obtiendrons $\text{cov}(y) = D_{1,Nd}\Gamma D_{1,Nd}$

A première vue l'évaluation de la fonction de vraisemblance n'est pas simple. nous pouvons en effet l'écrire sous la forme générale

$$L_v(\phi, \theta) = (2\pi)^{-Nd/2} \left(\prod_{t=1}^{Nd} \sigma_t^2 \det \Gamma \right)^{-1/2} \exp \left\{ -\frac{1}{2} y^T D_{1,Nd}^{-1} \Gamma^{-1} D_{1,Nd}^{-1} y \right\} \quad (3.3.14)$$

d'où

$$\log l(\beta, y) = -Nd/2 \log 2\pi - 1/2 \log \text{Det} \Gamma - N/2 \sum_{t=1}^d \log \sigma_i^2 - 1/2 y^T D_{1,Nd}^{-1} \Gamma^{-1} D_{1,Nd}^{-1} y. \quad (3.3.15)$$

En explicitant l'équation normale par rapport à σ_i^2 et en le remplaçant dans l'équation (3.3.15) ce qui fourni l'estimateur du maximum de vraisemblance.

Il faut optimiser la fonction L_v par des routines numériques qui nécessitent l'évaluation de celle-ci au point courant. Il semble donc qu'il faut inverser une matrice $Nd \times Nd$ ce qui requiert un nombre d'opérations de l'ordre de $(Nd)^3$. Une autre difficulté mis à part que le nombre d'opérations croit très vite avec la taille de l'observation est que l'expression de la matrice de covariance Γ est très compliquée. Dans ce qui suit nous avons choisit de détailler un algorithme efficient (dû à Vecchia, (1985a)) pour l'estimation des paramètres par la méthode du maximum de vraisemblance non conditionnelle.

4. Estimation par la méthode du Maximum de Vraisemblance d'un modèle PARMA

4.1. Le modèle et sa fonction de vraisemblance :

Considérons une série chronologique y_1, y_2, \dots, y_{Nd} supposée obtenue à partir du processus périodiquement corrélé d -périodique $\{y_{t(i,\tau)}, i=1, \dots, d, \tau \in \mathbb{Z}\}$ où $t(i,\tau) = i + d\tau$;

$1 \leq i \leq d$ est l'index cumulatif du temps, τ : l'index de l'année ($\tau = 0, \pm 1, \pm 2, \dots$) et i l'indice de saison ($i = 1, 2, \dots, d$).

Nous supposons sans perte de généralité que $E(y_{t(i,\tau)}) = 0 \quad \forall \tau \in \mathbb{Z}, \forall i \in \{1, \dots, d\}$ (le cas où $\{y_{t(i,\tau)}\}$ n'est pas centré se ramène au cas précédent voir plus haut.)

Nous rappelons que le processus $\{y_{t(i,\tau)}\}$ est dite suivre un modèle $ARMA_d(p_t, q_t)$ si l'équation suivante est vérifiée pour tout i et τ ,

$$y_{t(i,\tau)} - \sum_{j=1}^{p_t} \phi_{ij} y_{t(i,\tau)-j} = \varepsilon_{t(i,\tau)} - \sum_{j=1}^{q_t} \theta_{ij} \varepsilon_{t(i,\tau)-j} \quad (3.4.1)$$

Nous supposons aussi que :

- Les paramètres p_t et q_t sont constants. Le cas où p_t et q_t sont périodiques de période d est traité par Vecchia (1983).
- Les innovations $\{\varepsilon_{t(i,\tau)}\}$ sont non corrélées de moyenne nulle et de variance périodique dans le temps. C'est à dire : $E(\varepsilon_{t(i,\tau)}) = 0$ et $var(\varepsilon_{t(i,\tau)}) = \sigma_i^2$ et comme nous allons considérer la fonction de vraisemblance de (3.4.1) donc nous admettons que les $\{\varepsilon_{t(i,\tau)}\}$ sont normalement distribuées et que les données sont disponibles pour les années $\tau = 0, 1, \dots, N-1$.

Comme nous avons vu au chapitre 2, le modèle (3.4.1) peut s'exprimer en une forme équivalente ARMA multivariée.

$\Phi(L)y_\tau = \Theta(L)\varepsilon_\tau$ où $y_\tau = [y_{t(\tau,1)}, y_{t(\tau,2)}, \dots, y_{t(\tau,d)}]^T$, $\Phi(L)$ et $\Theta(L)$ sont des polynômes de matrices en L dont l'expression en détail est donnée dans le chapitre 2.

Remarque

- Pour donner une bonne estimation, il est nécessaire de supposer que le processus $\{y_{t(i,\tau)}\}$ est causal (stationnaire).
- Une méthode d'obtenir les estimateurs de maximum de vraisemblance est de convertir (3.4.1) à sa forme ARMA multivariée équivalente puis utiliser les résultats de Hilmer et Tiao, (1979). Mais le coût de calcul et de capacité de stockage nous incite à utiliser une autre approche.

La première étape consiste à écrire la fonction de vraisemblance dans la forme la plus simple à évaluer : soit le vecteur $y = [y_1, y_2, \dots, y_{Nd}]^T$. La fonction de vraisemblance de y est donnée par :

$$L(y) = (2\pi)^{-Nd/2} |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2} y^T \Sigma^{-1} y\right\}$$

où $\Sigma = cov(y)$ est la matrice de variance-covariance qui sera fonction des paramètres θ et ϕ .

Toutefois, la fonction donnée si dessus présente deux inconvénients majeurs : Le 1^{er} est que le calcul de Σ est une tâche difficile. La seconde est que la dimension de Σ augmente rapidement avec N ce qui rend le calcul de son déterminant et son inverse plus compliqué. Pour contourner ces problèmes Vecchia (1985a) a développé une technique efficace pour le calcul de la fonction de vraisemblance exacte dans le cas où p est pas nul et une approximation valide pour N suffisamment grand dans le cas où p n'est pas nul. Nous commençons tout d'abord par le cas où p est nul.

4.1.1. La fonction de vraisemblance quand $p = 0$

$p = 0$ (3.4.1) devient donc :

$$y_{t(i,\tau)} = \varepsilon_{t(i,\tau)} - \sum_{j=1}^q \theta_{ij} \varepsilon_{t(i,\tau)-j} \quad (3.4.2)$$

et considérons la construction suivante :

$$y_m = [y_{m+1}, y_{m+2}, \dots, y_{Nd}]^T$$

avec $0 \leq m \leq Nd - 1$

$$\varepsilon_m = [\varepsilon_{m+1}, \varepsilon_{m+2}, \dots, \varepsilon_{Nd}]^T$$

En explicitant chaque composante de y_m selon l'équation (3.4.2) nous aurons le système suivant

$$y_m = L_{\theta,m} \varepsilon_m - M_{\theta,m} \varepsilon_* \quad (3.4.3)$$

où

$$\varepsilon_* = [\varepsilon_{m+1-q}, \varepsilon_{m+2-q}, \dots, \varepsilon_m]^T$$

$L_{\theta,m}$ est une $(Nd - m) \times (Nd - m)$ matrice triangulaire inférieure définie par :

$$(L_{\theta,m})_{jk} = \begin{cases} 1 & j = k \\ 0 & j < k \\ -\theta_{m+j, j-k} & j > k \end{cases}$$

$M_{\theta,m}$ est une $(Nd - m) \times q$ matrice avec :

$$(M_{\theta,m})_{jk} = \begin{cases} \theta_{m+j, q+j-k} & i \leq q \\ 0 & i > q \quad \text{ou} \quad i < j \end{cases}$$

Proposition 3.4.1

La fonction de vraisemblance du vecteur y_m de (3.4.2) est donnée par :

$$\begin{aligned} L(\theta, \sigma / y_m) &= \left(\frac{1}{\sqrt{2\pi}} \right)^{(Nd-m)} |D_{m+1-q, Nd}|^{-1/2} |A_{\theta,m}|^{-1/2} \\ &\times \exp \left\{ -\frac{1}{2} \left[\hat{\varepsilon}_*^T D_{m+1-q, m}^{-1} \hat{\varepsilon}_* + \hat{\varepsilon}_m^T D_{m+1, Nd}^{-1} \hat{\varepsilon}_m \right] \right\} \end{aligned} \quad (3.4.4)$$

où :

$$D_{ij} = \text{diag} \{ \sigma^2(i), \dots, \sigma^2(j) \},$$

$$A_{\theta,m} = D_{m+1-q, m}^{-1} + F_{\theta,m}^T D_{m+1, Nd}^{-1} F_{\theta,m},$$

$$F_{\theta,m} = -L_{\theta,m}^{-1} M_{\theta,m}, \quad \hat{\varepsilon}_* = E(\varepsilon_* / y_m) \text{ et}$$

$$\hat{\varepsilon}_m = E(\varepsilon_m / y_m).$$

Preuve

Nous avons d'après (3.4.3) :

$$y_m = L_{\theta,m} \varepsilon_m - M_{\theta,m} \varepsilon_*$$

avec :

ε_m est un vecteur gaussien de moyenne nulle et de matrice de variance-covariance $D_{m+1, Nd}$,

ε_* est un vecteur gaussien de moyenne nulle et de matrice de variance-covariance $D_{m+1-q, m}$.

Comme ε_m et ε_* sont indépendants car $\{\varepsilon_{t(i, \tau)}\}$ sont indépendants. Par conséquent la densité conditionnelle de ε_m sachant ε_* est la même que la densité de ε_m .

$$f(\varepsilon_m / \varepsilon_*) = \left(1 / \sqrt{2\pi}\right)^{Nd-m} |D_{m+1, Nd}|^{-1/2} \exp\left\{-\frac{1}{2} \varepsilon_m^T D_{m+1, Nd}^{-1} \varepsilon_m\right\}. \quad (3.4.5)$$

D'autres part, à partir de (3.4.3) nous avons : $\varepsilon_m = L_{\theta, m}^{-1} (y_m + M_{\theta, m} \varepsilon_*)$

par conséquent le Jacobien de la transformation de $(\varepsilon_m / \varepsilon_*)$ vers (y_m / ε_*) est tel que tous les éléments de la diagonale valent 1. Il se réduira donc à 1.

Ainsi,

$$\begin{aligned} f(y_m / \varepsilon_*) &= f(\varepsilon_m(y_m) / \varepsilon_*) \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^{Nd-m} |D_{m+1, Nd}|^{-1/2} \exp\left\{-\frac{1}{2} (y_m + M_{\theta, m} \varepsilon_*)^T L_{\theta, m}^{T-1} D_{m+1, Nd}^{-1} L_{\theta, m}^{-1} (y_m + M_{\theta, m} \varepsilon_*)\right\} \end{aligned} \quad (3.4.6)$$

En multipliant $f(y_m / \varepsilon_*)$ par $f(\varepsilon_*)$, nous obtenons $f(y_m, \varepsilon_*)$, la densité conjointe de (y_m, ε_*) :

$$\begin{aligned} f(y_m, \varepsilon_*) &= \left(\frac{1}{\sqrt{2\pi}}\right)^q |D_{m+1-q, m}|^{-1/2} \exp\left\{-\frac{1}{2} \varepsilon_*^T D_{m+1-q, m}^{-1} \varepsilon_*\right\} \times \\ &\quad \left(\frac{1}{\sqrt{2\pi}}\right)^{NS-m} |D_{m+1, Nd}|^{-1/2} \exp\left\{-\frac{1}{2} (y_m + M_{\theta, m} \varepsilon_*)^T L_{\theta, m}^{T-1} D_{m+1, Nd}^{-1} L_{\theta, m}^{-1} (y_m + M_{\theta, m} \varepsilon_*)\right\} \end{aligned}$$

En utilisant le fait que :

$$|D_{m+1-q, m}| \cdot |D_{m+1, Nd}| = \begin{vmatrix} D_{m+1-q, m} & 0 \\ 0 & D_{m+1, Nd} \end{vmatrix} = |D_{m+1-q, Nd}|$$

et que $x'Ax + y'By = \begin{pmatrix} x \\ y \end{pmatrix}' \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$

nous aurons :

$$f(y_m, \varepsilon_*) = \left(\frac{1}{\sqrt{2\pi}}\right)^{Nd-m+q} |D_{m+1-q, m}|^{-1/2} \exp\left\{-\frac{1}{2} (\Gamma y_m - X \varepsilon_*)^T D_{m+1-q, Nd}^{-1} (\Gamma y_m - X \varepsilon_*)\right\} \quad (3.4.7)$$

où :

$$\Gamma = \begin{bmatrix} 0 \\ L_{\theta, m}^{-1} \end{bmatrix} \text{ et } X = \begin{bmatrix} I_q \\ -L_{\theta, m}^{-1} M_{\theta, m} \end{bmatrix};$$

L'exposant de $f(y_m, \varepsilon_*)$ peut s'écrire :

$$\begin{aligned} (\Gamma y_m - X \varepsilon_*)^T D_{m+1-q, Nd}^{-1} (\Gamma y_m - X \varepsilon_*) &= \\ &= \left(\Gamma y_m - X \varepsilon_* + X \hat{\varepsilon}_* - X \hat{\varepsilon}_*\right)^T D_{m+1-q, Nd}^{-1} \left(\Gamma y_m - X \varepsilon_* + X \hat{\varepsilon}_* - X \hat{\varepsilon}_*\right) \\ &= \left((\Gamma y_m - X \hat{\varepsilon}_*) - X(\varepsilon_* - \hat{\varepsilon}_*)\right)^T D_{m+1-q, Nd}^{-1} \left((\Gamma y_m - X \hat{\varepsilon}_*) - X(\varepsilon_* - \hat{\varepsilon}_*)\right) \\ &= (\Gamma y_m - X \hat{\varepsilon}_*)^T D_{m+1-q, Nd}^{-1} (\Gamma y_m - X \hat{\varepsilon}_*) + (\varepsilon_* - \hat{\varepsilon}_*)^T X^T X (\varepsilon_* - \hat{\varepsilon}_*) - \\ &\quad - 2(\varepsilon_* - \hat{\varepsilon}_*)^T X^T D_{m+1-q, Nd}^{-1} (\Gamma y_m - X \hat{\varepsilon}_*) \end{aligned}$$

où $\hat{\varepsilon}_*$ est choisit de manière à ce que : $-2(\varepsilon_* - \hat{\varepsilon}_*)^T X^T D_{m+1-q, Nd}^{-1} (\Gamma y_m - X \hat{\varepsilon}_*) = 0$

i.e. :

$$(\varepsilon_* - \hat{\varepsilon}_*)^T (X^T D_{m+1-q, Nd}^{-1} \Gamma y_m - X^T D_{m+1-q, Nd}^{-1} X \hat{\varepsilon}_*) = 0$$

Il suffit que

$$X^T D_{m+1-q, Nd}^{-1} \Gamma y_m - X^T D_{m+1-q, Nd}^{-1} X \hat{\varepsilon}_* = 0$$

d'où

$$\hat{\varepsilon}_* = (X^T D_{m+1-q, Nd}^{-1} X)^{-1} X^T D_{m+1-q, Nd}^{-1} \Gamma y_m \quad (3.4.8)$$

Finalement,

$$f(y_m, \varepsilon_*) =$$

$$\left(\frac{1}{\sqrt{2\pi}} \right)^{Nd-m} |D_{m+1-q, Nd}|^{-1/2} |X^T D_{m+1-q, Nd} X|^{-1/2} \exp \left\{ -\frac{1}{2} (\Gamma y_m - X \hat{\varepsilon}_*)^T D_{m+1-q, Nd}^{-1} (\Gamma y_m - X \hat{\varepsilon}_*) \right\} \times$$

$$\left(\frac{1}{\sqrt{2\pi}} \right)^q |X^T D_{m+1-q, Nd} X|^{-1/2} \exp \left\{ -\frac{1}{2} (\varepsilon_* - \hat{\varepsilon}_*)^T D_{m+1-q, Nd}^{-1} (\varepsilon_* - \hat{\varepsilon}_*) \right\}$$

$$f(y_m, \varepsilon_*) = f(y_m) \times f(\varepsilon_* / y_m)$$

Avec :

$$f(y_m) = \left(\frac{1}{\sqrt{2\pi}} \right)^{NS-m} |D_{m+1-q, NS}|^{-1/2} |X^T D_{m+1-q, NS} X|^{-1/2} \exp \left\{ -\frac{1}{2} (\Gamma y_m - X \hat{\varepsilon}_*)^T D_{m+1-q, NS}^{-1} (\Gamma y_m - X \hat{\varepsilon}_*) \right\}$$

est la fonction de vraisemblance de θ et de σ^2 qui relativement au vecteur fixé y_m .

Après manipulations Algébriques, la fonction de vraisemblance devient :

$$L(\theta, \sigma / y_m) = \left(\frac{1}{\sqrt{2\pi}} \right)^{Nd-m} |D_{m+1-q, Nd}|^{-1/2} |A_{\theta, m}|^{-1/2} \exp \left\{ -\frac{1}{2} \hat{\varepsilon}_*^T D_{m+1-q, Nd}^{-1} \hat{\varepsilon}_* - \frac{1}{2} \hat{\varepsilon}_m^T D_{m+1, Nd}^{-1} \hat{\varepsilon}_m \right\} \quad (3.4.9)$$

Avec :

$$A_{\theta, m} = A_{m+1-q, m}^{-1} + F_{\theta, m}^T D_{m+1, Nd}^{-1} F_{\theta, m}$$

$$F_{\theta, m} = -L_{\theta, m}^{-1} M_{\theta, m}$$

$$\hat{\varepsilon}_m = E(\varepsilon_m / y_m) = L_{\theta, m}^{-1} y_m + L_{\theta, m}^{-1} M_{\theta, m} \hat{\varepsilon}_*$$

$$\hat{\varepsilon}_* = E(\varepsilon_* / y_m) = A_{\theta, m}^{-1} F_{\theta, m}^T D_{m+1, Nd}^{-1} L_{\theta, m}^{-1} y_m \quad (3.4.10)$$

Remarques

La fonction de vraisemblance donnée ci-dessus présente les avantages suivants :

Les formes incluant D_{ij}^{-1} sont simples à calculer car D_{ij} est une matrice diagonale.

$L_{\theta, m}$ est une matrice triangulaire inférieure avec éléments diagonaux égal à 1, alors l'inverse de $F_{\theta, m}$ est facilement calculable de manière récursive.

$A_{\theta, m}$ est une matrice $q \times q$ définie positive et sa dimension n'augmente pas avec N et son déterminant peut être calculer par la décomposition de Cholesky.

Pour $\hat{\varepsilon}_*$ donné les éléments $\hat{\varepsilon}_m$ peuvent être obtenus de manière récursive par :

$$\hat{\varepsilon}_{m+j} = y_{m+j} + \sum_{k=1}^q \theta_{m+j,k} \hat{\varepsilon}_{m+j-k} \quad j = 1, 2, \dots, Nd - m \quad (3.4.11)$$

Les valeurs initiales, $\{\theta_k\}$ seront déduites à partir des équations de Yule-Walker périodiques.

La seule difficulté rencontrée est la détermination de $\hat{\varepsilon}_*$ où

$$\hat{\varepsilon}_* = A_{\theta,m}^{-1} F_{\theta,m}^T D_{m+1, Nd}^{-1} L_{\theta,m}^{-1} y_m \quad (3.4.12)$$

mais à partir de (3.4.3) il est évident que $E(\varepsilon_m / y_m, \varepsilon_* = 0) = L_{\theta,m}^{-1} y_m$ donc $L_{\theta,m}^{-1} y_m$ peut être déduite à partir de (3.4.12) avec $\hat{\varepsilon}_* = 0$ ainsi il reste à calculer l'inverse de $A_{\theta,m}$.

4.1.2. Approximation de la fonction de vraisemblance pour $p > 0$

Pour $p > 0$ nous ne pouvons calculer la fonction de vraisemblance exacte en utilisant la méthode précédente, cependant une transformation adéquate permet d'obtenir une bonne approximation de celle-ci.

Pour $m = 0$ la quantité $L(\theta, \sigma / y_0)$ est la fonction de vraisemblance exacte du modèle. Quand nous rajoutons une composante autorégressive AR au modèle précédent (i.e. $p > 0$) nous devons calculer la fonction de covariance d'un modèle $ARMA_d(p, q)$ qui présente une tâche difficile, surtout lorsque p ou d deviennent grands. Donc quand $p > 0$ la fonction de vraisemblance exacte doit être remplacée par la fonction de vraisemblance des $(Nd-p)$ dernières observations. Tout en fixant les p premières.

Soit alors la transformation suivante :

$$w_{p+j} = y_{p+j} - \sum_{k=1}^p \phi_{p+j,k} y_{p+j-k} \quad j = 1, 2, \dots, Nd - p \quad (3.4.13)$$

et soit $w_p = [w_{p+1}, w_{p+2}, \dots, w_{Nd}]^T$. Cette transformation permet de passer à une série $\{w_{t(i,\tau)}\}$ d'un modèle $MA_d(q)$ en considérant y_1, y_2, \dots, y_p fixes. La transformation de y_p à w_p a un Jacobien égale à 1 (car : $w_p = L_{\phi,p} y_p - M_{\phi,p} y_*$ en gardant les mêmes notation et $|L_{\phi,p}|^{-1} = 1$).

Or w_p suit un modèle $MA_d(q)$ périodique en commençant à l'instant $p+1$ donc d'après la section précédente le logarithme de la fonction de vraisemblance de w_p est donnée par :

$$\begin{aligned} -2 \text{Log} [L(\phi, \theta, \sigma / w_p)] = \\ (Nd - p) \text{Log} 2\pi + \text{Log} |D_{p+1-q, Nd}| + \text{Log} |A_{\theta,p}| + \hat{\varepsilon}_*^T D_{p+1-q,p}^{-1} \hat{\varepsilon}_* + \hat{\varepsilon}_p^T D_{p+1, Nd}^{-1} \hat{\varepsilon}_p \end{aligned} \quad (3.4.14)$$

où

$D_{ij}, A_{\theta,p}, F_{\theta,p}$ sont données selon la formule (3.4.10).

$$\hat{\varepsilon}_* = \left[\hat{\varepsilon}_{p+1-q}, \hat{\varepsilon}_{p+2-q}, \dots, \hat{\varepsilon}_p \right]^T = A_{\theta,p}^{-1} F_{\theta,p}^T D_{p+1, Nd}^{-1} L_{\theta,p}^{-1} w_p \quad (3.4.15)$$

et

$$\hat{\varepsilon}_{p+j} = w_{p+j} + \sum_{k=1}^q \theta_{p+j,k} \hat{\varepsilon}_{p+j-k} \quad j = 1, \dots, Nd - p \quad (3.4.16)$$

4.2. Estimation du maximum de vraisemblance

Les estimateurs des paramètres : ϕ, θ, σ obtenus en minimisant (3.4.14) sont appelés les estimateurs de maximum de vraisemblance. Lorsque p est strictement positif, la vraisemblance est basée sur seulement les $(Nd - p)$ observations : $y_{p+1}, y_{p+2}, \dots, y_{NS}$ pour y_1, y_2, \dots, y_p fixés. Nous pouvons vérifier facilement que si $q = 0$ l'estimateur obtenu est équivalent aux estimateurs de Yule-Walker saisonnier (voir Pagano 1978). Cependant, tout le long de ce travail nous admettons que q est strictement positif au quel cas la minimisation directe de (3.4.14) est pratiquement infaisable. L'algorithme donné ci-dessous, dû à Vecchia (1985a), permet de converger en un temps considérablement petit, par rapport à celui obtenu par minimisation directe de (3.4.14).

Le principe de l'algorithme est le suivant :

Nous commençons par des estimations initiales $\hat{\phi}_0$ et $\hat{\theta}_0$ obtenues à partir de la méthodes des moments.

Nous considérons que $\hat{\varepsilon}_* = 0$ et nous calculons $\hat{\varepsilon}_m$ selon la formule récursive (3.4.16).

A partir de $\hat{\varepsilon}_m$ et $\hat{\varepsilon}_*$ Nous estimons $\hat{\sigma}_i^2$ puis nous minimisons la quantité $-2 \text{Log} \left[L(\hat{\phi}, \hat{\theta}, \hat{\sigma} / w_p) \right]$.

Nous recalculons $\hat{\varepsilon}_m$ et $\hat{\varepsilon}_*$ selon (3.5.16) et (3.5.15) et nous réitérons cette procédure jusqu'à ce que le critère d'arrêt choisit (convergence) soit vérifié.

4.3. Algorithme

Etape (0)

Calculer les estimateurs initiaux $\tilde{\phi}_0, \tilde{\theta}_0$ de ϕ et θ en utilisant les équations de Yule-Walker périodiques.

Etape (1)

Poser $\hat{\varepsilon}_* = 0$ et calculer $\hat{\varepsilon}_m$ par la formule

$$\hat{\varepsilon}_{p+j} = w_{p+j} + \sum_{k=1}^q \theta_k (p+j) \hat{\varepsilon}_{p+j-k} \quad j=1, \dots, Nd-p$$

Etape (2)

Soit :

K_1 le plus grand entier satisfaisant : $K_1 d < p + 1 - q$

K_2 le plus grand entier satisfaisant : $K_2 d < p + 1 + d$

Alors on calcule les estimateurs des variances résiduelles

$$\hat{\sigma}_i^2 = (N - K_2)^{-1} \sum_{n=1}^{N-K_1} \hat{\varepsilon}_{K_1 d + i + (n-1)d}$$

où $\hat{\varepsilon}_j = 0$ quand il existe j vérifiant : $K_1 d < j \leq p - q$

Etape (3)

Utiliser une routine d'optimisation non linéaire pour déterminer $\hat{\phi}$ et $\hat{\theta}$ minimisant $-2 \text{Log} \left[L(\phi, \theta, \hat{\sigma} / w_p) \right]$

Etape (4)

Calculer $\hat{\varepsilon}_m$ et $\hat{\varepsilon}_*$ à partir des formules

$$\hat{\varepsilon}_* = \left[\varepsilon_{p+1-q}, \varepsilon_{p+2-q}, \dots, \varepsilon_p \right]^T = A_{\theta,p}^{-1} F_{\theta,p}^T D_{p+1, Nd}^{-1} L_{\theta,p}^{-1} w_p$$

$$\hat{\varepsilon}_{p+j} = w_{p+j} + \sum_{k=1}^q \theta_{p+j,k} \varepsilon_{p+j-k} \quad j = 1, \dots, Nd-p$$

Répéter les étapes (2), (3), (4) jusqu'à obtenir une convergence à la précision voulue.

Fin de l'algorithme.

5. Application numérique

Nous avons simulé 1000 séries de longueur variant de 80 à 200, supposées obtenues à partir de modèles PARMA différents. Nous avons fait varier les paramètres de façon à ce que les propriétés de causalité et d'inversibilité soient vérifiées. Pour chaque modèle nous avons appliqué les différentes méthodes exposées dans ce chapitre, l'objet de cette section n'est pas de faire une comparaison complète des différentes méthodes mais plutôt l'obtention d'une sorte de synthèse. Les sorties de simulations seront les moyennes et les écart-types, sur le nombre de répétitions, des estimations produites. Notons que la mesure de dispersion RMSE (*root minimum square error*) est aussi utilisée. Les résultats seront donnés sous forme de tableaux.

Modèle PARMA₂(1,0)

Paramètre	N=80	MM	RMSE	MCC	RMSE	MV	RMSE
ϕ_1	0.7	0.6890	0.3264	0.8167	1.0252	0.6891	1.1235
ϕ_2	0.2	0.1966	0.0700	0.2119	1.2587	0.1971	1.3548
σ_1^2	2	2.0844	0.4675	2.1322	0.6808	1.9907	0.8769
σ_2^2	10	9.4821	2.2147	10.9865	1.9527	9.7025	1.8972

Tableau 3.1

Paramètre	N=100	MM	RMSE	MCC	RMSE	MV	RMSE
ϕ_1	0.7	0.6811	0.3255	0.7867	1.0252	0.7105	1.3584
ϕ_2	0.2	0.2060	0.0553	0.2133	1.2587	0.2110	1.2357
σ_1^2	2	2.0688	0.3440	1.8897	0.4985	1.8793	0.9658
σ_2^2	10	9.3283	1.7596	11.8972	2.3168	11.0284	1.2547

Tableau 3.2

Paramètre	N=120	MM	RMSE	MCC	RMSE	MV	RMSE
ϕ_1	0.7	0.6882	0.2845	0.6841	0.8792	0.7204	1.5492
ϕ_2	0.2	0.1976	0.0516	0.1917	1.3197	0.1987	0.5847
σ_1^2	2	2.0763	0.3762	2.2195	0.5784	2.1209	2.5478
σ_2^2	10	9.2924	1.7332	9.8623	2.2968	10.9854	1.2587

Tableau 3.2

Paramètre	N=140	MM	RMSE	MCC	RMSE	MV	RMSE
ϕ_1	0.7	0.6995	0.2327	0.7267	1.0252	0.7125	1.1124
ϕ_2	0.2	0.2019	0.0513	0.2333	1.2587	0.2198	2.3687
σ_1^2	2	2.0821	0.3219	1.9562	0.4985	2.1138	0.7895
σ_2^2	10	9.5036	1.7868	10.7271	2.3168	9.2146	1.5746

Tableau 3.4

Paramètre	N=160	MM	RMSE	MCC	RMSE	MV	RMSE
ϕ_1	0.7	0.7147	0.2481	0.7167	1.1568	0.6901	0.9874
ϕ_2	0.2	0.2003	0.0532	0.1819	1.4587	0.1897	1.2547
σ_1^2	2	2.0977	0.3609	1.9897	0.6689	2.0193	0.9874
σ_2^2	10	9.5964	1.5287	9.8473	2.1058	11.2359	1.5789

Tableau 3.5

Paramètre	N=180	MM	RMSE	MCC	RMSE	MV	RMSE
ϕ_1	0.7	0.7111	0.2280	0.7155	1.1973	0.7018	0.5879
ϕ_2	0.2	0.1984	0.0451	0.2033	1.1895	0.1987	1.4789
σ_1^2	2	2.0590	0.2929	1.8328	0.3429	2.0489	0.8587
σ_2^2	10	9.5543	1.4343	10.1254	2.0095	10.6589	1.5874

Tableau 3.6

Paramètre	N=200	MM	RMSE	MCC	RMSE	MV	RMSE
ϕ_1	0.7	0.6974	0.2164	0.7055	1.2568	0.7047	0.8974
ϕ_2	0.2	0.1970	0.0411	0.2013	1.4495	0.2078	1.2578
σ_1^2	2	2.0638	0.2887	2.0358	0.7658	2.1476	1.2547
σ_2^2	10	9.5932	1.4791	10.0254	1.9785	10.3587	1.6587

Tableau 3.7

Modèle PARMA₃(1,0)

Paramètre	N=100	MM	RMSE	MCC	RMSE	MV	RMSE
ϕ_1	-0.9	-0.7997	0.8579	-0.8512	1.2578	-0.879	0.8974
ϕ_2	0.4	0.5178	1.2547	0.3754	2.0487	0.3697	1.2587
ϕ_3	0.2	0.2314	0.9587	0.1857	1.2591	0.1954	0.8433
σ_1^2	1	1.2314	1.6587	1.3258	1.3354	1.2586	1.2541
σ_2^2	1	0.9874	0.9874	1.0578	0.7299	1.2581	1.6582
σ_3^2	1	1.4978	1.5872	0.8795	1.2873	1.1247	0.6987

Tableau 3.8

Paramètre	N=150	MM	RMSE	MCC	RMSE	MV	RMSE
ϕ_1	-0.9	-0.8069	1.4725	-0.8236	1.5551	-1.0489	1.0213
ϕ_2	0.4	0.3748	0.9587	0.4388	0.6471	0.4211	0.8821
ϕ_3	0.2	0.1872	1.2547	0.2313	2.0104	0.2143	1.2546
σ_1^2	1	1.0971	1.1694	1.1123	1.8721	0.8814	0.5268
σ_2^2	1	0.8974	1.0028	0.9885	0.6587	1.2456	1.2549
σ_3^2	1	1.2038	1.6675	1.2255	1.2894	1.2583	1.2450

Tableau 3.9

Paramètre	N=200	MM	RMSE	MCC	RMSE	MV	RMSE
ϕ_1	-0.9	-0.8476	1.2451	-0.8615	1.3584	-0.8712	1.0214
ϕ_2	0.4	0.3874	0.2457	0.4285	0.8495	0.4089	2.0351
ϕ_3	0.2	0.2208	0.8741	0.1954	0.63214	0.2198	1.2116
σ_1^2	1	1.1354	1.2971	1.3584	0.6584	0.9845	1.3254
σ_2^2	1	1.0192	0.7854	0.9524	1.5498	1.2548	1.4512
σ_3^2	1	0.9854	1.3549	1.3258	2.2584	1.0219	1.6782

Tableau 3.10

Modèle PARMA₄(1,0)

Paramètre	N=100	MM	RMSE	MV	RMSE
ϕ_1	1.5	1.5901	0.8945	1.5613	1.5483
ϕ_2	-0.5	-0.4301	0.5234	-0.4751	0.2581
ϕ_3	1.2	1.3109	1.2548	1.1547	0.9658
ϕ_4	0.7	0.8457	1.3695	0.6512	0.4567
σ_1^2	1	1.2354	0.6985	1.3291	1.3254
σ_2^2	1	0.9842	0.8952	1.0257	0.9618
σ_3^2	1	1.3512	0.7583	1.2961	0.9944
σ_4^2	1	1.1024	1.9281	1.5522	1.2541

Tableau 3.10

Paramètre	N=160	MM	RMSE	MV	RMSE
ϕ_1	1.5	1.5647	0.3651	1.5393	1.4984
ϕ_2	-0.5	-0.4423	1.2546	-0.4725	0.3584
ϕ_3	1.2	1.2235	1.3651	1.2604	1.0354
ϕ_4	0.7	0.6299	1.9662	0.8022	0.5612
σ_1^2	1	1.2540	0.3558	1.2332	1.1259
σ_2^2	1	0.8882	2.0210	1.3035	1.3954
σ_3^2	1	1.0941	0.9654	1.2230	0.7787
σ_4^2	1	1.2005	0.9841	0.9621	1.3540

Tableau 3.11

Paramètre	N=200	MM	RMSE	MV	RMSE
ϕ_1	1.5	1.5539	0.4521	1.5206	1.5471
ϕ_2	-0.5	-0.5367	1.3240	-0.4895	0.4028
ϕ_3	1.2	1.3654	1.5421	1.1851	1.0025
ϕ_4	0.7	0.6654	1.7548	0.7312	0.5247
σ_1^2	1	1.0234	0.4823	1.1129	1.3675
σ_2^2	1	0.9239	1.7869	1.0058	1.0025
σ_3^2	1	1.1124	0.8477	1.2597	1.1657
σ_4^2	1	1.2254	1.3254	1.0025	0.9871

Tableau 3.12

Modèle PARMA₇(1,0)

Paramètre	N=100	MM	RMSE	MCC	RMSE
ϕ_1	1.5	1.4371	1.2347	1.5382	0.9618
ϕ_2	-0.5	-0.599	0.4325	-0.4596	0.6548
ϕ_3	1.2	1.1631	1.3574	1.0936	1.2937
ϕ_4	0.7	0.6661	0.6574	0.7695	0.6999
ϕ_5	0.8	0.7780	1.2547	0.8462	0.4598
ϕ_6	-1.4	-1.3859	0.3259	-1.4819	1.5362
ϕ_7	0.9	0.8747	1.2586	0.7607	0.3254
σ_1^2	1	1.1231	1.3252	1.1149	1.3254
σ_2^2	1	0.91247	1.2433	1.1954	0.8954
σ_3^2	1	0.9952	0.5421	1.1364	0.7410
σ_4^2	1	1.2033	1.0953	0.8651	1.2361
σ_5^2	1	1.1157	1.0257	0.9248	0.9248
σ_6^2	1	1.1987	0.5621	1.0354	1.3594
σ_7^2	1	0.8247	0.3658	1.1941	0.2357

Tableau 3.13

Paramètre	N=150	MM	RMSE	MCC	RMSE
ϕ_1	1.5	1.4201	1.1325	1.5213	0.8136
ϕ_2	-0.5	-0.5176	0.5943	-0.4406	0.5349
ϕ_3	1.2	1.1120	1.0325	1.1359	1.1291
ϕ_4	0.7	0.7414	0.5684	0.7381	0.7329
ϕ_5	0.8	0.7503	1.1378	0.8214	0.4421
ϕ_6	-1.4	-1.3530	0.4451	-1.4521	1.2541
ϕ_7	0.9	0.9157	1.1189	0.8311	0.1235
σ_1^2	1	1.0231	1.3126	0.9124	1.4567
σ_2^2	1	0.9519	1.2791	1.2351	0.1328
σ_3^2	1	0.9803	0.6349	1.0951	0.6731
σ_4^2	1	1.1046	1.3569	0.9234	1.1143
σ_5^2	1	1.0152	1.3258	1.1369	0.9523
σ_6^2	1	0.9401	0.8534	0.9561	1.1327
σ_7^2	1	0.8829	0.4329	1.1349	0.4521

Tableau 3.14

Paramètre	N=200	MM	RMSE	MCC	RMSE
ϕ_1	1.5	1.3560	1.2319	1.4823	0.2314
ϕ_2	-0.5	-0.4972	0.4217	-0.4841	0.4875
ϕ_3	1.2	1.1876	1.1349	1.1920	0.9542
ϕ_4	0.7	0.6965	0.4531	0.7127	1.1382
ϕ_5	0.8	0.7836	1.2999	0.8299	0.6543
ϕ_6	-1.4	-1.4115	0.4178	-1.3913	1.1634
ϕ_7	0.9	0.9064	1.1336	0.8812	0.4369
σ_1^2	1	1.1354	1.5421	0.9561	0.2841
σ_2^2	1	0.9821	1.1954	1.0384	1.5483
σ_3^2	1	1.0654	0.3584	1.1002	0.5621
σ_4^2	1	0.9217	1.1254	0.9575	1.5671
σ_5^2	1	0.9521	0.9541	1.0654	1.0238
σ_6^2	1	1.0354	1.0747	0.9234	0.9842
σ_7^2	1	0.9290	0.6543	0.9217	0.6214

Tableau 3.15

Modèle PARMA₂(2,0)

Paramètre	N=100	MM	RMSE	MCC	RMSE
ϕ_{11}	0.2	0.2028	0.2345	0.1843	0.3548
ϕ_{12}	-0.5	-0.6443	1.4628	-0.4526	1.2358
ϕ_{21}	0.9	0.9590	0.4943	0.8649	0.5373
ϕ_{22}	0.7	0.8185	1.2373	0.6663	1.3693
σ_1^2	5	4.3611	2.1546	5.3617	1.9872
σ_2^2	9	8.7234	2.1591	9.2376	2.3671

Tableau 3.16

Paramètre	N=150	MM	RMSE	MCC	RMSE
ϕ_{11}	0.2	0.2606	0.3254	0.2248	0.6541
ϕ_{12}	-0.5	-0.6229	1.3467	-0.5213	1.1254
ϕ_{21}	0.9	0.9346	0.6543	0.8741	0.6123
ϕ_{22}	0.7	0.6521	1.1137	0.7329	1.3325
σ_1^2	5	4.9009	2.2464	5.1395	1.1472
σ_2^2	9	8.7823	1.9871	8.9124	2.5481

Tableau 3.17

Paramètre	N=200	MM	RMSE	MCC	RMSE
ϕ_{11}	0.2	0.2230	0.5421	0.2136	0.5984
ϕ_{12}	-0.5	-0.5695	1.2254	-0.4987	1.2354
ϕ_{21}	0.9	0.8641	0.7412	0.9124	0.4587
ϕ_{22}	0.7	0.7456	1.2315	0.7171	1.1235
σ_1^2	5	4.8245	2.2354	4.9841	1.2135
σ_2^2	9	9.1443	2.3298	9.1037	2.3211

Tableau 3.18

Modèle PARMA₄(2,0)

Paramètre	N=100	MM	RMSE	MCC	RMSE
ϕ_{11}	-0.2	-0.2234	0.5487	-0.1824	0.5584
ϕ_{12}	0.35	0.3045	0.6548	0.3845	0.6458
ϕ_{21}	1.2	1.1013	0.9452	1.7123	0.8456
ϕ_{22}	-0.27	-0.2758	0.6548	-0.2952	1.2145
ϕ_{31}	-0.3	-0.2315	0.6397	-0.3356	0.6515
ϕ_{32}	0.54	0.5023	0.9654	0.5553	0.8852
ϕ_{41}	1.7	1.7360	1.0314	1.6679	0.8547
ϕ_{42}	-0.7	-0.6549	0.8745	-0.7327	0.9542
σ_1^2	1	1.2354	0.8957	0.9871	1.2478
σ_2^2	1	0.9017	0.5364	1.1125	0.7154

Tableau 3.19

Paramètre	N=160	MM	RMSE	MCC	RMSE
ϕ_{11}	-0.2	-0.1732	0.6548	-0.2254	0.5245
ϕ_{12}	0.35	0.2958	0.7253	0.3692	0.4523
ϕ_{21}	1.2	1.1965	0.6354	1.2153	0.8266
ϕ_{22}	-0.27	-0.2854	0.8521	-0.2992	1.1471
ϕ_{31}	-0.3	-0.3125	0.8470	-0.3421	0.6218
ϕ_{32}	0.54	0.5328	0.8173	0.5427	0.7854
ϕ_{41}	1.7	1.7215	1.1258	1.6926	0.8699
ϕ_{42}	-0.7	-0.6821	1.2401	-0.7164	0.8590
σ_1^2	1	1.1116	0.9075	0.9754	1.2112
σ_2^2	1	1.1254	0.5940	1.0325	0.7177

Tableau 3.20

Paramètre	N=200	MM	RMSE	MCC	RMSE
ϕ_{11}	-0.2	-0.2129	0.6678	-0.2025	0.6654
ϕ_{12}	0.35	0.3325	0.7324	0.3421	0.7412
ϕ_{21}	1.2	1.1748	0.8721	1.9628	0.4520
ϕ_{22}	-0.27	-0.2701	0.7732	-0.2552	0.8453
ϕ_{31}	-0.3	-0.3256	0.6231	-0.2716	0.7124
ϕ_{32}	0.54	0.5377	0.9541	0.5232	0.7326
ϕ_{41}	1.7	1.6810	1.1304	1.7321	0.9527
ϕ_{42}	-0.7	-0.7241	0.9229	-0.7125	0.8241
σ_1^2	1	1.0654	0.7548	1.0547	1.0235
σ_2^2	1	0.9813	0.8523	0.9650	0.8529

Tableau 3.21

Modèle PARMA₂(0,1)

Paramètre	N=100	MCC	RMSE	MVC	RMSE
θ_1	-0.4	-0.3221	0.8405	-0.4321	0.9845
θ_2	0.6	0.5587	0.8395	0.6325	0.9542
σ_1^2	1	1.2499	1.0245	0.9123	1.0245
σ_2^2	1	0.8759	1.2254	0.9542	1.2345

Tableau 3.22

Paramètre	N=150	MCC	RMSE	MVC	RMSE
θ_1	-0.4	-0.4421	0.9202	-0.3703	0.9548
θ_2	0.6	0.5621	0.9195	0.6245	1.0254
σ_1^2	1	1.2053	1.2304	1.1094	1.1123
σ_2^2	1	0.9245	1.2034	0.9555	1.2547

Tableau 3.23

Paramètre	N=200	MCC	RMSE	MVC	RMSE
θ_1	-0.4	-0.3878	0.9548	-0.4121	0.9054
θ_2	0.6	0.5821	0.9437	0.6025	0.9054
σ_1^2	1	1.0749	1.0234	0.9123	1.2304
σ_2^2	1	0.9230	1.1254	0.9842	1.2230

Tableau 3.24

Modèle PARMA₂(3,0)

Paramètre	N=100	Estimateur MCC	RMSE
ϕ_{11}	-0.9	-0.7952	1.0235
ϕ_{12}	0.4	0.3524	0.3145
ϕ_{13}	0.2	0.1952	0.5216
ϕ_{21}	-0.6	-0.5584	0.9854
ϕ_{22}	0.35	0.3354	0.4631
ϕ_{23}	-0.3	-0.2421	0.3467
σ_1^2	1	1.0394	1.0235
σ_2^2	1	0.9234	1.5241

Tableau 3.25

Paramètre	N=150	Estimateur MCC	RMSE
ϕ_{11}	-0.9	-0.8312	1.2354
ϕ_{12}	0.4	0.4256	0.4521
ϕ_{13}	0.2	0.2134	0.4325
ϕ_{21}	-0.6	-0.5421	0.8792
ϕ_{22}	0.35	0.3745	0.5214
ϕ_{23}	-0.3	-0.2847	0.2358
σ_1^2	1	0.9743	0.8546
σ_2^2	1	0.9432	1.4321

Tableau 3.26

Paramètre	N=200	Estimateur MCC	RMSE
ϕ_{11}	-0.9	-0.8654	0.9542
ϕ_{12}	0.4	0.3910	0.5511
ϕ_{13}	0.2	0.1923	0.5629
ϕ_{21}	-0.6	-0.6321	0.6624
ϕ_{22}	0.35	0.3214	0.2534
ϕ_{23}	-0.3	-0.2951	0.2754
σ_1^2	1	1.0654	0.9235
σ_2^2	1	0.9921	1.2104

Tableau 3.27

Modèle PARMA₂(1,1)

Paramètre	N=100	MCC	RMSE
ϕ_{11}	0.8	0.7135	1.2345
θ_{11}	0.3	0.2654	0.6548
ϕ_{21}	0.6	0.5548	0.7459
θ_{21}	0.5	0.5231	0.9854
σ_1^2	1	1.2340	1.0245
σ_2^2	1	0.8742	1.3254

Tableau 3.28

Paramètre	N=100	MCC	RMSE
ϕ_{11}	0.8	0.7465	1.1136
θ_{11}	0.3	0.3325	0.5632
ϕ_{21}	0.6	0.6231	0.6845
θ_{21}	0.5	0.4879	0.6789
σ_1^2	1	1.1254	0.9874
σ_2^2	1	0.9254	1.1354

Tableau 3.29

Paramètre	N=100	MCC	RMSE
ϕ_{11}	0.8	0.7732	1.2149
θ_{11}	0.3	0.3127	0.6548
ϕ_{21}	0.6	0.6065	0.5617
θ_{21}	0.5	0.4946	0.4987
σ_1^2	1	0.9548	1.0254
σ_2^2	1	1.0897	1.2154

Tableau 3.30

2^{ème} Partie

Estimation en-ligne

Chapitre 4

Estimation en-ligne dans les modèles ARMA classiques

1. Introduction

Le problème de l'estimation des paramètres de modèles de séries chronologiques a été largement traité, tant dans la littérature des statisticiens que celle des ingénieurs. Les techniques d'estimation peuvent être classées en deux catégories générales, selon que la taille de la série est fixe ou non : Les méthodes hors-ligne et les méthodes en-ligne. La première concerne le cas où la série sous-jacente est de taille fixe, la deuxième apparaît quand les données sont progressivement disponibles.

Les méthodes d'estimation hors-ligne basées essentiellement, sur la méthode du *pseudo-maximum de vraisemblance* ou sur la méthode des moindres carrés, s'avèrent trop lourdes voir impossibles lorsqu'elles sont appliquées à des données en-ligne. Car elles nécessitent à l'introduction de chaque nouvelle donnée de refaire complètement la procédure

d'estimation, ce qui est très coûteux en terme de complexité temporelle et spatiale. L'introduction, par les ingénieurs, des méthodes d'estimation récurrentes portant le nom *d'identification récurrente* a permis de soulever ce problème.

2. Méthodes d'erreur de prévision minimale (RPEM)

Parmi les nombreux algorithmes récurrents qui existent, nous avons choisi une famille d'algorithmes issue des méthodes d'estimation hors-ligne. Le critère de base pour les algorithmes de cette classe est la minimisation de l'erreur de prévision moyenne, exprimée sous forme empirique. Dans le cas gaussien, cette classe nommée '*Recursive Predictor Error Method*' (RPEM) conduit à la même fonction critère que le *maximum de vraisemblance conditionnelle*, et donne des estimateurs dont les qualités sont équivalentes à celui-ci. Nous commençons d'abord par traiter le plus ancien et le plus populaire des algorithmes récurrents. Il s'agit de l'algorithme des moindres carrés récursif (*Recursive Least Square*, RLS). Cette méthode est d'usage courant en contrôle adaptatif et en traitement de signal, l'un de ses avantages, du fait que la fonction critère à optimiser est quadratique par rapport aux paramètres, est que les estimateurs qu'elle produise sont obtenus de manière exacte. Cependant, l'handicaps majeur de cette méthode est qu'elle se restreint uniquement aux modèles, linéaires, ne comportant pas de composante moyenne mobile. En effet, en présence d'une composante moyenne mobile la fonction critère ne serait plus quadratique des paramètres et l'estimateur ne peut être obtenu analytiquement. Ainsi en opérant quelques approximations, l'optimisation numérique, au moyen des méthodes stochastiques de Gauss-Newton, fournit l'algorithme appelé *Recursive Maximum Likelihood* (RML). L'avantage de cet algorithme, bien qu'il donne des estimations calculées approximativement, est qu'il peut être appliqué à des situations plus larges (modèles ARMA, ARMAX, PARMA ...). De plus pour le cas des modèles autorégressifs purs l'algorithme RML se confond avec l'algorithme RLS.

2.1. Principes

Le principe de l'estimation récurrente est le suivant, étant donnée une série chronologique issue d'un processus stochastique du second ordre $\{y_t, t \in Z\}$, dont la taille n'est pas fixée à l'avance. Notons $y^t = \{y_1, y_2, \dots, y_t\}$ la série disponible à l'instant t . Supposons que cette série est générée à partir d'un modèle de série chronologique, soit ARMA(p, q). Les paramètres du modèle seront notés par le vecteur β et l'estimateur récurrent, $\hat{\beta}_t$, de β est fonction de y^t .

Notons que toute méthode d'estimation récurrente se contraint par la condition suivante : **Les complexités temporelles et spatiales doivent être indépendantes de la taille de l'observation** (en d'autres termes, l'espace mémoire et le temps de calcul ne croissent pas en même temps que t). Pour répondre à cette contrainte l'idée de base sera de condenser, dans une quantité auxiliaire S_t de dimension fixe (et dépendante du modèle), l'information, concernant β , contenue dans la série y^t . L'estimateur $\hat{\beta}_t$ et la quantité S_t seront alors calculés selon un algorithme ayant la structure suivante :

$$\begin{cases} \hat{\beta}_t = F(\hat{\beta}_{t-1}, S_t, y_t) \\ S_t = H(S_{t-1}, \hat{\beta}_{t-1}, y_t) \end{cases} \quad (4.2.1)$$

où y_t est l'observation de la série à l'instant t , F et H sont des fonctions dépendantes de la méthode utilisée et du modèle choisit. L'estimation $\hat{\beta}_t$ est fonction de $\hat{\beta}_{t-1}$, S_t , et de y_t . Donc à chaque instant t , il ne sera nécessaire de stocker que les quantités $\{\hat{\beta}_t, S_t\}$

2.2. Critères

Notons que toute méthode récurrente part d'une fonction objectif à réaliser, passe par le choix d'une structure de réalisation et de procédures d'optimisation, et ce traduit par un algorithme précis pour la mise en œuvre effective sur les données. Plusieurs critères peuvent être donc intéressants à optimiser (Approche Bayésienne, Maximum de vraisemblance,...). Mais souvent, le critère retenu pour optimiser le choix des paramètres dans les algorithmes récurrents est la minimisation de l'erreur quadratique moyenne, à savoir :

$$\beta^* = \underset{\beta \in M}{\text{Arg min}} \{V(\beta) = E(\hat{\varepsilon}_t^2(\beta))\} \quad (4.2.2)$$

où M est le référentiel des paramètres et $\hat{\varepsilon}_t(\beta)$ est l'erreur de prévision obtenue en approchant l'état courant (y_t) par sa prévision, notée $\hat{y}_{t/t-1}(\beta)$, basée sur le passé du processus sous étude. Ce critère théorique définit le but à atteindre par les algorithmes d'identification récurrente. Cependant, le critère de l'erreur quadratique moyenne cité ci-dessus peut faire l'objet d'un algorithme d'optimisation numérique, à condition que l'on sache obtenir la fonction critère $V(\beta)$ et ses propriétés caractéristiques (le gradient, la hessienne). Mais cela nécessite la connaissance a priori des propriétés statistiques au second ordre des processus traités. A ce moment là l'identification du paramètre β ne se fera pas directement à partir des observations. Afin de pouvoir estimer le paramètre du modèle par traitement direct des données, (nous supposons avoir N observations nous remplaçons donc le critère $V(\beta)$ par un critère calculable effectivement à l'aide des données. Une façon naturelle de le faire est de remplacer l'espérance théorique dans (4.2.2) par son estimation empirique. En d'autres termes il revient à minimiser le critère suivant

$$V_N(\beta) = \frac{1}{N} \sum_{i=1}^N \varepsilon_i^2(\beta) \quad (4.2.3)$$

Cette approche permet de construire une suite d'estimateurs $\hat{\beta}_t$ du paramètre vectoriel, permettant d'approcher de plus en plus la valeur β^* . La question sera donc d'étudier le comportement de $\hat{\beta}_t$ quand t tend vers l'infini, et notamment la convergence vers la solution β^* . L'ergodisme permet d'espérer un tel résultat dans la mesure où $V_N(\beta)$ tend vers $V(\beta)$ quand N tend vers l'infini.

Dans ce qui suit nous utiliserons le critère des moindres carrés pour la construction d'algorithmes récurrents. Avant de passer au schéma général de l'algorithme, nous allons montrer comment calculer, de manière récursive, la prévision ($\hat{y}_{t/t-1}(\beta)$) de y_t , faite à l'instant $t-1$, pour l'horizon 1. Cette étape, comme nous allons voir ultérieurement, est indispensable pour la construction de l'algorithme récurrent.

2.3. Calcul récursif de l'erreur de prévision

Zahaf, (1998) a pris comme exemple le modèle ARMA(1,1) pour décrire la méthode RPEM, cependant nous avons préféré le modèle ARMA(p,q) général. Nous allons donc montrer comment évaluer la prévision $\hat{y}_{t/t-1}(\beta)$ ainsi que l'erreur $\hat{\varepsilon}_t(\beta)$ à l'instant t .

Considérons le modèle ARMA(p,q) suivant

$$y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \dots - \phi_p y_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (4.2.4)$$

où $\{\varepsilon_t, t \in Z\}$ est un bruit blanc de variance σ^2 et $\beta = (\phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q)^T$ est le vecteur des paramètres du modèle.

Alors la prévision, $\hat{y}_{t/t-1}(\beta)$, de y_t faite à l'instant $t-1$, sera définie par

$$\hat{y}_{t/t-1}(\beta) = y_t - \hat{\varepsilon}_t(\beta) \quad (4.2.5)$$

où $\hat{\varepsilon}_t(\beta)$ est une approximation de l'erreur ε_t qui s'obtient par récurrence sur la base de y_{t-k} , $k = 1, \dots, p$ et $\hat{\varepsilon}_{t-l}(\beta)$, $l = 1, \dots, q$. Et ce, selon l'équation suivante

$$\hat{\varepsilon}_t = y_t - \phi_1 y_{t-1} - \dots - \phi_p y_{t-p} + \theta_1 \hat{\varepsilon}_{t-1} - \dots - \theta_q \hat{\varepsilon}_{t-q} \quad (4.2.6)$$

Par suite la prévision $\hat{y}_{t/t-1}(\beta)$ peut se mettre sous la forme

$$\hat{y}_{t/t-1}(\beta) = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} - \theta_1 \hat{\varepsilon}_{t-1} - \theta_2 \hat{\varepsilon}_{t-2} - \dots - \theta_q \hat{\varepsilon}_{t-q} \quad (4.2.7)$$

Il est nécessaire d'avoir des valeurs initiales $y_0, y_{-1}, \dots, y_{1-p}$, $\hat{\varepsilon}_0, \hat{\varepsilon}_{-1}, \dots, \hat{\varepsilon}_{1-q}$ mais l'effet de celles-ci décroît en supposant que les paramètres vérifient les conditions d'inversibilité, nous pouvons alors les choisir arbitrairement sans affecter la prévision.

Notons qu'à l'instant t , les observations y_{t-k} étant connues et ce, pour $k = 1, \dots, p$.

Pour avoir une formule de récurrence exprimant la prévision $\hat{y}_{t/t-1}(\beta)$ nous remplaçons les termes $\hat{\varepsilon}_t(\beta)$ par $y_t - \hat{y}_{t/t-1}(\beta)$ dans l'équation (4.2.7), après manipulation il vient que

$$\hat{y}_{t/t-1}(\beta) - \theta_1 \hat{y}_{t-1/t-2}(\beta) - \dots - \theta_q \hat{y}_{t-q/t-q-1}(\beta) = (\phi_1 - \theta_1) y_{t-1} + \dots + (\phi_h - \theta_h) y_{t-h} \quad (4.2.8)$$

où $h = \max(p, q)$. Pour simplifier les valeurs initiales $\hat{y}_{1/0}, \hat{y}_{0/-1}, \dots, \hat{y}_{1-q/-q}$ sont mises égales à zéro.

Nous avons donc vu comment obtenir, de manière récurrente et pour des valeurs des paramètres connus a priori, la prévision $\hat{y}_{t/t-1}(\beta)$ et par conséquent l'erreur de prévision.

3. L'algorithme des moindres carrés récursif (RLS)

Cet algorithme est dérivé de la méthode des moindres carrés usuelle qui s'applique aux modèles autorégressifs, donné sous une forme récursive. Considérons le modèle autorégressif d'ordre p suivant (AR(p))

$$y_t - \phi_1 y_{t-1} - \dots - \phi_p y_{t-p} = \varepsilon_t \quad (4.3.1)$$

où $\{\varepsilon_t, t \in \mathbb{Z}\}$ est un bruit blanc et $\beta = (\phi_1, \phi_2, \dots, \phi_p)^T$ est le vecteur des paramètres à estimer à partir de la donnée de y_t pour $t = 1, \dots, N$. Introduisons maintenant le vecteur suivant

$\varphi_t^T = (y_{t-1}, y_{t-2}, \dots, y_{t-p})$, alors (4.3.1) peut s'écrire comme

$$y_t = \beta^T \varphi_t + \varepsilon_t \quad (4.3.2)$$

Ce modèle décrit la variable y_t comme combinaison linéaire des composantes de φ_t plus un bruit blanc. Une bonne prévision de y_t basée sur son passé et que l'on note $\hat{y}_{t/t-1}(\beta)$, est par conséquent donnée

$$\hat{y}_{t/t-1}(\beta) = \beta^T \varphi_t \quad (4.3.3)$$

Le critère (4.2.3) devient

$$V_N(\beta) = \frac{1}{N} \sum_{i=1}^N [y_i - \beta^T \varphi_i]^2 \quad (4.3.4)$$

La minimisation de la fonction critère qui, grâce à la linéarité de la prévision par rapport à β , est quadratique, peut être obtenue analytiquement. Ce qui permet d'estimer β , à travers

$$\hat{\beta}_N = \left[\sum_{t=1}^N \varphi_t \varphi_t^T \right]^{-1} \sum_{t=1}^N \varphi_t y_t \quad (4.3.5)$$

sous l'hypothèse que l'inverse de $\sum_{t=1}^N \varphi_t \varphi_t^T$ existe. $\hat{\beta}_N$ est alors défini comme l'estimateur des moindres carrés de β .

Mis à part la complexité de calcul pour l'obtention de $\hat{\beta}_N$ (inversion d'une matrice d'ordre N), l'inconvénient de cette méthode directe (ou encore 'bloc') est double :

- Nécessité de stockage des données.
- Impossibilité de traitement en-ligne, en effet, la prise en compte de nouvelles données nécessite de refaire complètement les calculs en traitant un bloc de taille encore supérieure.

Notons que le résultat (4.3.5) peut être obtenu de façon récursive, ceci se fera par mise à jour permettant le calcul de $\hat{\beta}_{t+1}$ à partir de $\hat{\beta}_t$ et d'une nouvelle observation à l'instant

$t+1$, tout en respectant la contrainte de départ (4.3.1). Notons par $\bar{R}_t = \sum_{k=1}^t \varphi_k \varphi_k^T$

Alors à partir de (4.3.5) il vient que

$$\sum_{k=1}^t \varphi_k y_k = \bar{R}_{t-1} \hat{\beta}_{t-1}$$

à partir de la définition de \bar{R}_t , il s'en suit que

$$\bar{R}_{t-1} = \bar{R}_t - \varphi_t \varphi_t^T$$

donc

$$\begin{aligned} \hat{\beta}_t &= \bar{R}_t^{-1} \left[\sum_{k=1}^{t-1} \varphi_k \varphi_k^T + \varphi_t y_t \right] \\ &= \bar{R}_t^{-1} \left[\bar{R}_{t-1} \hat{\beta}_{t-1} + \varphi_t y_t \right] \\ &= \bar{R}_t^{-1} \left[\bar{R}_t \hat{\beta}_{t-1} + \varphi_t (-\varphi_t^T \hat{\beta}_{t-1} + y_t) \right] \end{aligned}$$

d'où

$$\hat{\beta}_t = \hat{\beta}_{t-1} + \bar{R}_t^{-1} \varphi_t [y_t - \varphi_t^T \hat{\beta}_{t-1}] \quad (4.3.6)$$

et

$$\bar{R}_t = \bar{R}_{t-1} + \varphi_t \varphi_t^T \quad (4.3.7)$$

Souvent, il est préférable de travailler avec $R_t = \frac{1}{t} \bar{R}_t$, à partir de (4.3.7) nous trouvons facilement

$$R_t = \frac{1}{t} [\bar{R}_{t-1} + \varphi_t \varphi_t^T] = \frac{t-1}{t} R_{t-1} + \frac{1}{t} \varphi_t \varphi_t^T$$

d'où

$$R_t = R_{t-1} + \frac{1}{t} [\varphi_t \varphi_t^T - R_{t-1}] \quad (4.3.8)$$

Les expressions précédentes se résument alors comme suit

$$\begin{cases} \hat{\beta}_t = \hat{\beta}_{t-1} + \frac{1}{t} R_t^{-1} \varphi_t [y_t - \hat{\beta}_{t-1}^T \varphi_t] \\ R_t = R_{t-1} + \frac{1}{t} [\varphi_t \varphi_t^T - R_{t-1}] \end{cases} \quad (4.3.9)$$

Les équations (4.3.9) possèdent théoriquement la forme souhaitée d'un algorithme récurrent. A l'instant t , seulement les valeurs de $\hat{\beta}_t$, R_t , y_t et φ_t vont être stockées en mémoire. En comparant à (4.2.1), S_t correspond à R_t , φ_t .

Remarque

L'algorithme décrit par les équations (4.3.9) n'est pas encore bien exploitable du point de vue numérique, car, à chaque itération, nous devons inverser une matrice de taille : la dimension du vecteur des paramètres. La complexité de l'algorithme sera donc de l'ordre de $O(p^3)$. Une forme équivalente et plus simple à évaluer consiste à introduire

$$P_t = \bar{R}_t^{-1} = \frac{1}{t} R_t^{-1}$$

et de calculer P_t directement à partir des équations de récurrence. Cela peut être accompli au moyen du résultat connu sous le nom du lemme d'inversion matricielle, qui s'énonce comme suit

Lemme 4.3.1

Soit A une matrice et b un vecteur de dimensions compatible de manière à ce que la somme $A + bb^T$ existe. Alors

$$\left[A + bb^T \right]^{-1} = A^{-1} - \frac{A^{-1}bb^T A^{-1}}{\left[1 + b^T A^{-1}b \right]} \quad (4.3.10)$$

Appliquons (4.3.10) à (4.3.6) et (4.3.7) avec $A = P_t^{-1}$, $b = \varphi_t$. Il s'en suit alors que

$$P_t = \left[P_t^{-1} + \varphi_t \varphi_t^T \right]^{-1}$$

d'où

$$P_t = P_{t-1} - \frac{P_{t-1} \varphi_t \varphi_t^T P_{t-1}}{1 + \varphi_t^T P_{t-1} \varphi_t} \quad (4.3.11)$$

L'avantage de (4.3.11) par rapport à (4.3.8), est la réduction de la dimension de la matrice à inverser (inversion d'un scalaire au lieu de l'inversion d'une matrice, R_t d'ordre $\dim \beta$). De même

$$P_t \varphi_t = P_{t-1} \varphi_t - \frac{P_{t-1} \varphi_t \varphi_t^T P_{t-1} \varphi_t}{1 + \varphi_t^T P_{t-1} \varphi_t}$$

donc

$$P_t \varphi_t = \frac{P_{t-1} \varphi_t}{1 + \varphi_t^T P_{t-1} \varphi_t} \quad (4.3.12)$$

Ainsi nous obtenons la forme complète d'implémentation numérique de l'algorithme ci-dessous.

$$\begin{cases} \varepsilon_t = y_t - \hat{\beta}_{t-1} \varphi_t \\ L_t = \frac{P_{t-1} \varphi_t}{1 + \varphi_t^T P_{t-1} \varphi_t} \\ \hat{\beta}_t = \hat{\beta}_{t-1} + L_t \varepsilon_t \\ P_t = P_{t-1} - L_t \varphi_t^T P_{t-1} \end{cases} \quad (4.3.13)$$

où $\varphi_t^T = (y_{t-1}, y_{t-2}, \dots, y_{t-p})$.

Ces dernières équations définissent l'algorithme des moindres carrés récursif (RLS). Cet algorithme est l'un des plus utilisés dans les méthodes d'identification numérique grâce à sa portabilité et sa robustesse. Les opérations de calcul intervenant dans les équations (4.3.13) nécessitent p^2 multiplications, et une mémorisation de vecteurs de taille p et d'une matrice de taille p^2 . La complexité temporelle et spatiale sera donc de l'ordre de $O(p^2)$.

Nous discutons, dans le chapitre implémentation, en détails les aspects algorithmiques de l'algorithme RLS.

4. Description de l'algorithme RML (*Recursive Maximum Likelihood*)

4.1. Préliminaires

Cet algorithme, issu de la méthode RPEM, est une généralisation de l'algorithme RLS (voir Michaut, (1991)) dans le cas de modèles à composante moyenne mobile. Il a été introduit par Aström, (1972) et Soderström, (1973b) et est appelé *Recursive Maximum Likelihood* (RML) en raison du lien entre les fonctions critères, alors que l'estimateur obtenu n'est pas du maximum de vraisemblance. En effet, la fonction critère ne sera pas quadratique des paramètres et l'estimateur ne peut être obtenu analytiquement. Cette méthode a été appliquée sur des données réelles et des données simulées avec un grand succès. Le critère choisit est le même (à une constante multiplicative près) que celui de (4.2.3), en effet il revient à choisir l'estimateur qui minimise la fonction critère suivante :

$$V_N(\beta) = \frac{1}{2} \sum_{i=1}^N \varepsilon_i^2(\beta)$$

Cette quantité est la même que le log négatif du maximum de vraisemblance conditionnelle gaussienne.

4.2. Optimisation stochastique

Soit $\hat{\beta}_{t-1}$ notre estimateur à l'instant $t-1$, c'est à dire celui qui minimise la quantité $V_{t-1}(\beta)$. Nous voulons obtenir l'estimateur $\hat{\beta}_t$ qui (approximativement) minimise la fonction critère à l'instant t . Développons par la formule de Taylor à l'ordre 2 la fonction $V_t(\beta)$ autour de $\hat{\beta}_{t-1}$. Nous aurons alors

$$V_t(\beta) = V_t(\hat{\beta}_{t-1}) + \frac{\partial V_t(\hat{\beta}_{t-1})}{\partial \beta^T} [\beta - \hat{\beta}_{t-1}] + \frac{1}{2} [\beta - \hat{\beta}_{t-1}]^T \frac{\partial^2 V_t(\hat{\beta}_{t-1})}{\partial \beta \partial \beta^T} [\beta - \hat{\beta}_{t-1}] + o(|\beta - \hat{\beta}_{t-1}|^2), \quad (4.4.1)$$

où $(\partial./\partial \beta^T)$ note la différentiation par rapport à β et $o(x)$ note une fonction telle que $\frac{o(x)}{|x|} \rightarrow 0$ quand $|x| \rightarrow 0$. Pour minimiser la fonction $V_t(\beta)$, nous dérivons le membre droit de l'équation (4.4.1) par rapport à β , l'expression obtenue est mise égale à zéro. Nous obtenons donc

$$\hat{\beta}_t = \hat{\beta}_{t-1} - \left[\frac{\partial^2 V_t(\hat{\beta}_{t-1})}{\partial \beta \partial \beta^T} \right]^{-1} \left[\frac{\partial V_t(\hat{\beta}_{t-1})}{\partial \beta^T} \right]^T + o(|\hat{\beta}_t - \hat{\beta}_{t-1}|) \quad (4.4.2)$$

Cette dernière équation, semblant déterminer le vecteur des paramètres β de façon récursive, n'est pas tout à fait complète, nous allons à cet effet faire quelques approximations pour l'évaluer. Notons $\psi_t(\beta)$ l'opposé de la dérivée de $\hat{\beta}_t(\beta)$ par rapport à β

$$\psi_t(\beta) = \left[-\frac{\partial \hat{\varepsilon}_t(\beta)}{\partial \beta^T} \right]^T \quad (4.4.3)$$

à partir de (4.2.3) nous avons

$$\left[\frac{\partial V_t(\beta)}{\partial \beta^T} \right]^T = -\sum_{k=1}^t \psi_k(\beta) \hat{\varepsilon}_k(\beta) = \left[\frac{\partial V_{t-1}(\beta)}{\partial \beta^T} \right]^T - \psi_t(\beta) \hat{\varepsilon}_t(\beta) \quad (4.4.4)$$

et par différentiation une fois de plus,

$$\frac{\partial^2 V_t(\beta)}{\partial \beta \partial \beta^T} = \frac{\partial^2 V_{t-1}(\beta)}{\partial \beta \partial \beta^T} + \psi_t(\beta) \psi_t^T(\beta) + \frac{\partial^2 \varepsilon_t(\beta)}{\partial \beta \partial \beta^T} \hat{\varepsilon}_t(\beta) \quad (4.4.5)$$

Pour évaluer (4.4.2), un certain nombre d'approximations est introduit, dans l'ordre.

1. Nous supposons que $\hat{\beta}_t$ se trouve dans un voisinage de $\hat{\beta}_{t-1}$. Cette hypothèse devient de plus en plus plausible lorsque t s'accroît. Cette supposition permet d'écrire les deux approximations suivantes :

$$\text{Négligeons } o(|\hat{\beta}_t - \hat{\beta}_{t-1}|) \text{ dans (4.4.2)} \quad (4.4.6)$$

Et à cause de la continuité des dérivées secondes, prenons

$$\frac{\partial^2 V_t(\hat{\beta}_t)}{\partial \beta \partial \beta^T} = \frac{\partial^2 V_t(\hat{\beta}_{t-1})}{\partial \beta \partial \beta^T} \quad (4.4.7)$$

2. Ensuite nous supposons que $\hat{\beta}_{t-1}$ est bien l'estimateur optimal à l'instant $t-1$, tel que

$$\frac{\partial V_{t-1}(\hat{\beta}_{t-1})}{\partial \beta^T} = 0 \quad (4.4.8)$$

Enfin posons

$$\frac{\partial^2 \hat{\varepsilon}_t(\hat{\beta}_{t-1})}{\partial \beta \partial \beta^T} \hat{\varepsilon}_t(\hat{\beta}_{t-1}) \approx 0 \quad (4.4.9)$$

La raison de cette dernière approximation est la suivante. Prés de la valeur β^* , $\{\hat{\varepsilon}_t(\beta)\}$ sera un bruit blanc, et ainsi nous pouvons approximativement considérer que $\hat{\varepsilon}_t(\beta)$ est de moyenne nulle et indépendant de ce qui c'est passé avant l'instant $t-1$. En particulier, il sera alors indépendant de $\left(\frac{\partial^2 \hat{\varepsilon}_t(\hat{\beta}_{t-1})}{\partial \beta \partial \beta^T} = -\frac{\partial^2 \hat{y}_{t-1}(\beta)}{\partial \beta \partial \beta^T} \right)$. L'espérance mathématique du membre de gauche de l'expression (4.4.9) est alors bien proche de 0, ainsi le dernier terme de (4.4.5) contribue beaucoup moins à $\left(\frac{\partial^2 V_t(\beta)}{\partial \beta \partial \beta^T} \right)$ que le second terme.

Sous les suppositions (4.4.9) et (4.4.7) insérées dans (4.4.5) nous pouvons évaluer approximativement la matrice dérivée seconde. Notons cette matrice par \bar{R}_t , alors nous avons

$$\bar{R}_t = \bar{R}_{t-1} + \psi_t(\hat{\beta}_{t-1}) \psi_t^T(\hat{\beta}_{t-1}) \quad (4.4.10)$$

En se servant de (4.4.8) et (4.4.4), nous aurons

$$\left[\frac{\partial V_t(\hat{\beta}_{t-1})}{\partial \beta} \right]^T = -\psi_t(\hat{\beta}_{t-1}) \hat{\varepsilon}_t(\hat{\beta}_{t-1}) \quad (4.4.11)$$

Utilisons cette dernière équation et l'approximation de (4.4.10) dans (4.4.2), nous aurons le système

$$\begin{cases} \bar{R}_t = \bar{R}_{t-1} + \psi_t(\hat{\beta}_{t-1}) \psi_t^T(\hat{\beta}_{t-1}) \\ \hat{\beta}_t = \hat{\beta}_{t-1} + \bar{R}_t^{-1} \psi_t(\hat{\beta}_{t-1}) \hat{\varepsilon}_t(\hat{\beta}_{t-1}) \end{cases} \quad (4.4.12)$$

Il reste maintenant, à discuter comment déterminer $\hat{\varepsilon}_t(\hat{\beta}_{t-1})$ et $\psi_t(\hat{\beta}_{t-1})$. A partir de (4.2.4), nous avons que

$$\psi_t^T(\beta) = -\frac{\partial \hat{\varepsilon}_t(\beta)}{\partial \beta} = \frac{\partial \hat{y}_{t/t-1}(\beta)}{\partial \beta^T} \quad (4.4.13)$$

à partir de (4.2.8), et par dérivation par rapport aux paramètres nous trouvons que

$$\left\{ \begin{array}{l} \frac{\partial \hat{y}_{t/t-1}(\beta)}{\partial \phi_1} - \theta_1 \frac{\partial \hat{y}_{t-1/t-2}(\beta)}{\partial \phi_1} - \theta_2 \frac{\partial \hat{y}_{t-2/t-3}(\beta)}{\partial \phi_1} - \dots - \theta_q \frac{\partial \hat{y}_{t-q/t-q-1}(\beta)}{\partial \phi_1} = y_{t-1} \\ \frac{\partial \hat{y}_{t/t-1}(\beta)}{\partial \phi_2} - \theta_1 \frac{\partial \hat{y}_{t-1/t-2}(\beta)}{\partial \phi_2} - \theta_2 \frac{\partial \hat{y}_{t-2/t-3}(\beta)}{\partial \phi_2} - \dots - \theta_q \frac{\partial \hat{y}_{t-q/t-q-1}(\beta)}{\partial \phi_2} = y_{t-2} \\ \\ \frac{\partial \hat{y}_{t/t-1}(\beta)}{\partial \phi_p} - \theta_1 \frac{\partial \hat{y}_{t-1/t-2}(\beta)}{\partial \phi_p} - \theta_2 \frac{\partial \hat{y}_{t-2/t-3}(\beta)}{\partial \phi_p} - \dots - \theta_q \frac{\partial \hat{y}_{t-q/t-q-1}(\beta)}{\partial \phi_p} = y_{t-p} \\ \\ \frac{\partial \hat{y}_{t/t-1}(\beta)}{\partial \theta_1} - \theta_1 \frac{\partial \hat{y}_{t-1/t-2}(\beta)}{\partial \theta_1} - \theta_2 \frac{\partial \hat{y}_{t-2/t-3}(\beta)}{\partial \theta_1} - \dots - \theta_q \frac{\partial \hat{y}_{t-q/t-q-1}(\beta)}{\partial \theta_1} = \hat{y}_{t-1/t-2}(\beta) - y_{t-1} \\ \frac{\partial \hat{y}_{t/t-1}(\beta)}{\partial \theta_2} - \theta_1 \frac{\partial \hat{y}_{t-1/t-2}(\beta)}{\partial \theta_2} - \theta_2 \frac{\partial \hat{y}_{t-2/t-3}(\beta)}{\partial \theta_2} - \dots - \theta_q \frac{\partial \hat{y}_{t-q/t-q-1}(\beta)}{\partial \theta_2} = \hat{y}_{t-2/t-3}(\beta) - y_{t-2} \\ \\ \frac{\partial \hat{y}_{t/t-1}(\beta)}{\partial \theta_q} - \theta_1 \frac{\partial \hat{y}_{t-1/t-2}(\beta)}{\partial \theta_q} - \theta_2 \frac{\partial \hat{y}_{t-2/t-3}(\beta)}{\partial \theta_q} - \dots - \theta_q \frac{\partial \hat{y}_{t-q/t-q-1}(\beta)}{\partial \theta_q} = \hat{y}_{t-q/t-q-1}(\beta) - y_{t-q} \end{array} \right. \quad (4.4.14)$$

Compte tenu du fait que

$$\psi_t^T(\beta) = -\frac{\partial \hat{\varepsilon}_t(\beta)}{\partial \beta^T} = \frac{\partial \hat{y}_{t/t-1}(\beta)}{\partial \beta^T} \quad (4.4.15)$$

nous aurons

$$\psi_t(\beta) - \theta_1 \psi_{t-1}(\beta) - \theta_2 \psi_{t-2}(\beta) - \dots - \theta_q \psi_{t-q}(\beta) = \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ \dots \\ y_{t-p} \\ -\hat{\varepsilon}_{t-1}(\beta) \\ -\hat{\varepsilon}_{t-2}(\beta) \\ \dots \\ -\hat{\varepsilon}_{t-q}(\beta) \end{pmatrix} \quad (4.4.16)$$

Ainsi, nous pouvons calculer $\hat{\varepsilon}_t(\hat{\beta}_{t-1})$ et $\psi_t(\hat{\beta}_{t-1})$ en résolvant (4.2.4), (4.2.8) et (4.4.16) respectivement. Ces équations sont des *filtres récurrents* (voir Ljung et Söderström (1983)) avec y comme entrée et ε , ψ comme sortie. Les coefficients du filtre sont déterminés par $\hat{\beta}_{t-1}$. Excepté pour le cas $\theta = 0$, les filtres ont une réponse impulsionnelle non tronquée. Ceci implique que toutes les données enregistrées y_s , $0 \leq s \leq t-1$, sont nécessaires pour déterminer $\hat{\varepsilon}_t(\hat{\beta}_{t-1})$ et $\psi_t(\hat{\beta}_{t-1})$, donc ces variables ne peuvent être calculées au sens de (4.2.1) en utilisant uniquement $\hat{\beta}_{t-1}$ et un vecteur de taille fixe.

4.3. Approximation de $\hat{\varepsilon}_t(\hat{\beta}_{t-1})$ et $\psi_t(\hat{\beta}_{t-1})$

Nous allons donner des approximations de $\hat{\varepsilon}_t(\hat{\beta}_{t-1})$ et $\psi_t(\hat{\beta}_{t-1})$ qui peuvent être calculées par récurrence au sens de (4.2.1). Une bonne approximation de $\hat{\varepsilon}_t(\hat{\beta}_{t-1})$ sera $\hat{\varepsilon}_t$, calculée suivant

$$\hat{\varepsilon}_t = y_t - \hat{y}_{t/t-1} \quad (4.4.17)$$

où la prévision $\hat{y}_{t/t-1}$ est calculée selon la formule (4.2.8) en remplaçant β par $\hat{\beta}_{t-1}$ avec $\hat{\beta}_{t-1} = (\hat{\phi}_1^{t-1}, \hat{\phi}_2^{t-1}, \dots, \hat{\phi}_p^{t-1}, \hat{\theta}_1^{t-1}, \hat{\theta}_2^{t-1}, \dots, \hat{\theta}_q^{t-1})^T$, c'est à dire que

$$\hat{y}_{t/t-1} = \hat{\theta}_1^{t-1} \hat{y}_{t-1/t-2} + \dots + \hat{\theta}_q^{t-1} \hat{y}_{t-q/t-q-1} + (-\hat{\theta}_1^{t-1} + \hat{\phi}_1^{t-1}) y_{t-1} + \dots + (-\hat{\theta}_h^{t-1} + \hat{\phi}_h^{t-1}) y_{t-h} \quad (4.4.18)$$

où $h = \max(p, q)$. Introduisons maintenant, le vecteur

$\varphi_t = (y_{t-1}, y_{t-2}, \dots, y_{t-p}, -\hat{\varepsilon}_{t-1}, -\hat{\varepsilon}_{t-2}, \dots, -\hat{\varepsilon}_{t-q})^T$. Alors nous pouvons écrire

$$\hat{\varepsilon}_t = y_t - \hat{\beta}_{t-1}^T \varphi_t \quad (4.4.19)$$

De même, une approximation de $\psi_t(\hat{\beta}_{t-1})$ sera alors, ψ_t calculée à partir de la formule suivante

$$\psi_t = \hat{\theta}_1^{t-1} \psi_{t-1} + \hat{\theta}_2^{t-1} \psi_{t-2} + \dots + \hat{\theta}_q^{t-1} \psi_{t-q} + \varphi_t \quad (4.4.20)$$

Donc $\hat{\varepsilon}_t$ et ψ_t seront calculées par récurrence à partir de (4.4.19) et (4.4.20). A l'instant t nous avons besoin de connaître seulement φ_t, ψ_t et $\hat{\beta}_{t-1}$. Finalement nous aurons l'algorithme suivant

$$\begin{cases} \varepsilon_t = y_t - \hat{\beta}_{t-1}^T \varphi_t \\ \hat{\beta}_t = \hat{\beta}_{t-1} + \bar{R}_t^{-1} \psi_t \varepsilon_t \\ \psi_{t+1} = \sum_{k=1}^q \hat{\theta}_k^t \psi_{t-k+1} + \varphi_{t+1} \\ \bar{R}_{t+1} = \bar{R}_t + \psi_{t+1} \psi_{t+1}^T \end{cases}$$

où $\varphi_t = (y_{t-1}, y_{t-2}, \dots, y_{t-p}, -\varepsilon_{t-1}, -\varepsilon_{t-2}, \dots, -\varepsilon_{t-q})^T$.

De la même manière que pour l'algorithme des moindres carrés récursif (RLS), et pour améliorer la performance numérique de l'algorithme RML nous remplaçons dans les équations de récurrence la matrice R_t par la matrice P_t , tel que :

$$P_t = \bar{R}_t^{-1} = \frac{1}{t} R_t^{-1}$$

Ensuite, nous calculons P_t directement à partir des équations de récurrence. Cela peut être accompli de deux manières différentes : Soit au moyen du lemme d'inversion matricielle ou bien sur la base de l'algorithme de factorisation U-D, de Biermann (1977), de la matrice P_t qui sera présentée au chapitre 6 (implémentation). Nous présentons pour le moment la forme de l'algorithme RML basée sur le lemme d'inversion matricielle.

$$\begin{cases} \hat{\varepsilon}_t = y_t - \hat{\beta}_{t-1}^T \psi_t \\ L_t = \frac{P_{t-1} \psi_t}{1 + \psi_t^T P_{t-1} \psi_t} \\ \hat{\beta}_t = \hat{\beta}_{t-1} + L_t \hat{\varepsilon}_t \\ P_t = P_{t-1} - L_t \psi_t^T P_{t-1} \end{cases} \quad (4.4.21)$$

Pour améliorer le comportement et la convergence de l'algorithme nous introduisons dans le schémas précédent la prévision de l'erreur notée $\bar{\varepsilon}_t$ et une suite, (γ) , de scalaires positifs tendant vers 0 telle que $(\sum \gamma_t)$ diverge. La structure des relations de récurrences est modifiée pour mieux faire apparaître le caractère récurrent de l'algorithme

$$\begin{cases} \hat{\varepsilon}_t = y_t - \hat{\beta}_t^T \bar{\varphi} \\ \hat{\beta}_t = \hat{\beta}_{t-1} + \gamma_t R_t^{-1} \psi_t \hat{\varepsilon}_t \\ \bar{\varepsilon}_t = y_t - \hat{\beta}_t^T \bar{\varphi} \\ \psi_{t+1} = \sum_{k=1}^q \hat{\theta}_{kt} \psi_{t-k+1} + \bar{\varphi}_{t+1} \\ R_{t+1} = R_t + \gamma_t [\psi_{t+1} \psi_{t+1}^T - R_t] \end{cases} \quad (4.4.22)$$

où $R_t = \gamma_t \bar{R}_t$, et $\bar{\varphi}_t = (y_t, \dots, y_{t-p+1}, -\bar{\varepsilon}_t, \dots, -\bar{\varepsilon}_{t-q+1})$.

Remarque:

- L'introduction de $\bar{\varepsilon}_t$ est justifiée par le fait que l'estimateur à l'instant t est meilleur que celui obtenu à l'instant $t-1$ et donc utiliser cette information supplémentaire permet d'améliorer la précision de l'algorithme. Dans le cas multivarié, Ljung et Söderström (1983) introduisent aussi le calcul récurrent de la matrice variance-covariance des erreurs, inutile par contre dans le cas univarié traité ici, mais qui sera utile dans le cas de modèle non stationnaire tels que les modèles périodiques qui seront abordés dans le chapitre suivant.
- La forme de l'algorithme décrit par les équations (4.4.22) peut être modifiée comme pour les équations (4.4.21) et ce pour éviter l'inversion matricielle dans chaque itération.

4.4. Estimation récurrente de la variance de l'innovation, σ_t^2

Nous choisissons comme estimateur, à l'instant t , de la variance du bruit blanc σ^2 , la variance empirique qui sera obtenue à partir de la relation suivante :

$$\hat{\sigma}_t^2 = 1/t \sum_{i=1}^t \hat{\varepsilon}_i^2 = (t-1)/t \hat{\sigma}_{t-1}^2 + (1/t) \hat{\varepsilon}_t^2. \quad (4.4.23)$$

Afin d'assurer la convergence de l'algorithme vers la solution optimale, un ensemble d'hypothèses sera nécessaire, nous discuterons cet aspect dans la section 6.

5. Algorithme de Mélard-Zahaf (RML_{MZ})**5.1 Méthode proposée**

Du point de vue théorique, sous un certain nombre de conditions théoriques, l'algorithme précédent donne des estimateurs convergents et asymptotiquement efficaces (voir section 6). Cependant, Mélard, (1989) a observé que les estimations successives de \bar{R}_t sont fort viables et que cela introduit des perturbations nuisibles dans l'estimation. En conservant le principe de cet algorithme, Zahaf, (1998) a proposé d'en améliorer la précision.

Zahaf, (1998) a proposé de remplacer la dernière équation de (4.3.14), qui n'est rien d'autre qu'une approximation de la hessienne ($\partial^2 V(\beta) / \partial \beta \partial \beta^T$), par l'espérance de celle-ci. En effet $\sigma^2 \bar{R}_t^{-1}$ représente en fait une approximation de la matrice covariance asymptotique $\Gamma(\beta^*)$ de l'estimateur $\hat{\beta}_t$ du maximum de vraisemblance (voir Zahaf (1998)).

Comme β^* est inconnu, Zahaf (1998) a proposé donc de remplacer la matrice de covariance par la matrice $\Gamma(\hat{\beta}_t)$. Si l'estimateur $\hat{\beta}_t$ converge vers β^* , ce qui sera établi plus tard, on peut espérer que $\Gamma(\hat{\beta}_t)$ converge vers $\Gamma(\beta^*)$. Or une bonne approximation de $\Gamma(\hat{\beta}_t)$ est l'inverse de la matrice d'information $F^{-1}(\hat{\beta}_t)$, où $F(\hat{\beta}_t)$ représente la matrice d'information de Fisher calculé en la valeur du paramètre $\hat{\beta}_t$. Nous calculerons à chaque

instant t la matrice d'information de Fisher $F(\hat{\beta}_{t-1})$ puis son inverse $F^{-1}(\hat{\beta}_{t-1})$ (noté F_t^{-1}) qui remplacera \bar{R}_t^{-1} dans l'algorithme donné précédemment :

$$\hat{\beta}_t = \hat{\beta}_{t-1} + \hat{\sigma}_t^{-2} F_t^{-1} \psi_t \hat{\varepsilon}_t \quad (4.5.1)$$

Le calcul par récurrence de \bar{R}_t , inutile, est supprimé et F_t^{-1} sera obtenue, par exemple, par la méthode suivante.

5.2. Calcul de la matrice d'information de Fisher pour un modèle ARMA(p, q)

L'objectif de ce paragraphe est de présenter un algorithme (dû à Klein et Mélard, (1989)) permettant le calcul des éléments (élément par élément) de la matrice d'information de Fisher. Cette matrice, étant très utile pour le calcul de la borne de Cramer-Rao, à fait l'objet d'une série de sujets de recherches afin de réduire, le plus possible, le temps de calcul.

Rappelons que (Cf. chapitre 1) la matrice d'information de Fisher F_N de l'estimateur du maximum de vraisemblance gaussien β_N , relative à une série y_1, y_2, \dots, y_N est donnée par

$$F_N = -E \left[\frac{\partial^2 \log l(\beta)}{\partial \beta \partial \beta^T} \right],$$

où $l(\beta)$ est la fonction de vraisemblance conditionnelle gaussienne.

Vu la structure particulière de celle ci (la matrice, symétrique, F est de Toeplitz) nous n'avons pas besoin alors de calculer tous ses éléments. Considérons alors la partition suivante citée par Godolphin et Unwin (1983) :

$$F = \begin{pmatrix} F_{\phi\phi} & F_{\phi\theta} \\ F_{\theta\phi} & F_{\theta\theta} \end{pmatrix}$$

Chacune des matrices $F_{\phi\phi}, F_{\phi\theta}, F_{\theta\theta}$ est Toeplitz. Rappelons qu'une matrice Toeplitz T (de taille $\ell \times \ell$) est de la forme

$$T = \begin{pmatrix} T_0 & T_{-1} & \dots & T_{-\ell+2} & T_{-\ell+1} \\ T_1 & T_0 & \ddots & \ddots & T_{-\ell+2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ T_{\ell-2} & \ddots & \ddots & \ddots & T_{-1} \\ T_{\ell-1} & T_{\ell-2} & \dots & T_1 & T_0 \end{pmatrix}$$

Pour N assez grand, nous avons

$$\log l(\beta) \approx -(2\sigma^2)^{-1} S(\beta) + cte,$$

où $S(\beta) = \sum_{t=1}^N \hat{\varepsilon}_t^2(\beta)$, $\hat{\varepsilon}_t(\beta) = \pi(L)y_t$ et $\pi(L) = \theta^{-1}(L)\phi(L)$. avec L est l'opérateur retard.

Soit F la matrice d'information limite nous avons alors $F_N = N.F$ (Cf. chapitre 1), puisque le

processus $\{y_t, t \in Z\}$ est stationnaire et $\hat{\varepsilon}_t(\beta)$ est non corrélé avec $\partial^2 \hat{\varepsilon}_t(\beta) / \partial \beta \partial \beta^T$ nous avons alors

$$NF = (2\sigma^2)^{-1} E \left[\frac{\partial^2 S(\beta)}{\partial \beta \partial \beta^T} \right] = N\sigma^{-2} E \left[\frac{\partial \hat{\varepsilon}_t(\beta)}{\partial \beta} \cdot \frac{\partial \hat{\varepsilon}_t(\beta)}{\partial \beta^T} \right], \quad (4.5.2)$$

Notons maintenant par $\varpi(L) = \phi^{-1}(L)\theta(L)$, où L est l'opérateur retard. Alors

$$\partial \hat{\varepsilon}_t(\beta) / \partial \beta^T = (\partial \pi(L) / \partial \beta^T) y_t = (\partial \pi(L) / \partial \beta^T) \varpi(L) \hat{\varepsilon}_t,$$

ou plus précisément

$$\begin{aligned} \frac{\partial \hat{\varepsilon}_t(\beta)}{\partial \phi_k} &= -L^k \theta^{-1}(L) y_t = -L^k \phi^{-1}(L) \hat{\varepsilon}_t(\beta), \\ \frac{\partial \hat{\varepsilon}_t(\beta)}{\partial \theta_j} &= L^j \theta^{-2}(L) \phi(L) y_t = L^j \theta^{-1}(L) \hat{\varepsilon}_t(\beta). \end{aligned}$$

Nous allons exploiter la forme Toepiltz de F (Akaike (1973)) et la symétrie de $F_{\phi\phi}, F_{\theta\theta}$ et $F_{\phi\theta}$, pour donner les éléments suivants :

$$\begin{aligned} (F_{\phi\phi})_{kj} &= (F_{\phi\phi})_{jk} = (F_{\phi\phi})_{j-k+1,1}; k=1,2,\dots,p; j=k,k+1,\dots,p \\ (F_{\theta\theta})_{kj} &= (F_{\theta\theta})_{jk} = (F_{\theta\theta})_{j-k+1,1}; k=1,2,\dots,q; j=k,k+1,\dots,q \\ (F_{\phi\theta})_{kj} &= (F_{\theta\phi})_{jk} = (F_{\phi\theta})_{j-k+1,1}; k=1,2,\dots,q; j=k,k+1,\dots,q \\ &= (F_{\theta\phi})_{1,k-j+1}; j=1,2,\dots,q; k=j,j+1,\dots,p. \end{aligned}$$

Seulement la première colonne de $F_{\phi\phi}$ et $F_{\theta\theta}$, ainsi que la première ligne et colonne de $F_{\phi\theta}$ ont besoin d'être calculés, soit $(q+p+(p+q-1))$ éléments en tout.

Il est connu (voir Whittle, (1953)) que le processus $\{z_t, t \in Z\}$ défini par $\theta(L)z_t = \phi(L)y_t$ a la même matrice d'information que le processus $\{y_t, t \in Z\}$. Or, suivant Pham (1986), considérons le processus autorégressif stationnaire $\{Z_t, t \in Z\}$ défini par

$$\phi(L)\theta(L)Z_t = \varepsilon_t.$$

ou encore

$$\Phi(L)Z_t = \varepsilon_t, \quad t \in Z$$

où $\Phi(L) = 1 - \sum_{j=1}^P \phi_j L^j$, $\phi_j = \sum_{k=0}^j -\phi_k \theta_{j-k}$ et $P = p+q$ avec la convention que $\phi_k, \theta_k = 0$ si $k > q$ ou $k < 0$. Puisque

$$\partial \varepsilon_t(\beta) / \partial \phi_k = -L^k \phi^{-1}(L) \varepsilon_t(\beta) = -L^k \theta(L) Z_t$$

et

$$\partial \varepsilon_t(\beta) / \partial \theta_j = L^j \theta^{-1}(L) \varepsilon_t(\beta) = L^j \phi(L) Z_t$$

les éléments de F peuvent donc être exprimés en fonction des autocovariances normalisées $\gamma_h = \sigma^{-2} \text{cov}(Z_t, Z_{t-h})$ du processus $\{Z_t, t \in Z\}$. Par exemple

$$\begin{aligned} (F_{\phi\phi})_{jk} &= \sigma^{-2} E[L^j \theta(L) Z_t L^k \theta(L) Z_t] \\ &= \sum_{s=0}^q \sum_{h=0}^q \theta_s \theta_h \gamma_{(s+j)-(h+k)}. \end{aligned}$$

Cette relation apparaît sous forme matricielle dans Pham (1986). Une implémentation rapide résulte de la considération des poids $\lambda_r = \sum_{u \in S_r} \theta_{u-r} \theta_u$, avec $S_r = \{r, r+1, \dots, q\}$ si $r \geq 0$, et $S_r = \{0, 1, \dots, q+r\}$ si $r \leq 0$. Alors

$$\begin{aligned} (F_{\phi\phi})_{jk} &= \sum_{r=-q}^q \gamma_{j-k+r} \lambda_r \\ &= (\gamma_{j-k} \lambda_0 + 2 \sum_{r=1}^q (\gamma_{j-k+r} - \gamma_{j-k-r}) \lambda_r). \end{aligned} \quad (4.5.3)$$

Puisque $\lambda_{-r} = \lambda_r$, le calcul de $\lambda_r = \theta_r + \theta_1 \theta_{r-1} + \dots + \theta_{q-r} \theta_q$ nécessite $q-r$ multiplications, et celui de (4.5.3), $q+1$, multiplications. Les p éléments $(i, 1)$, $i=1, 2, \dots, p$, de $F_{\phi\phi}$ peuvent être évalués en $(q+1)(q/2+p)$ multiplications.

De même

$$(F_{\theta\theta})_{jk} = (\gamma_{j-k} \mu_0 + 2 \sum_{r=1}^p (\gamma_{j-k+r} - \gamma_{j-k-r}) \mu_r),$$

et

$$(4.5.4)$$

$$\mu_r = \phi_r + \phi_1 \phi_{r-1} + \dots + \phi_{p-r} \phi_p$$

$$\begin{aligned} (F_{\theta\phi})_{jk} &= -\sigma^{-2} E[B^j \phi(B) Z_t B^k \theta(B) Z_t] \\ &= - \sum_{s=0}^p \sum_{h=0}^q \phi_s \theta_h \gamma_{(s+j)-(h+k)} \\ &= - \sum_{r=-q}^p \gamma_{j-k+r} v_r, \end{aligned} \quad (4.5.5)$$

$$\begin{aligned} \text{où } v_r &= \theta_{-r} + \phi_1 \theta_{1-r} + \dots + \phi_{r+q} \theta_q, \quad r = -q, -q+1, \dots, 0 \\ &= \phi_r + \phi_{r+1} \theta_1 + \dots + \phi_{r+q} \theta_q, \quad r = 1, 2, \dots, p-q \\ &= \phi_r + \phi_{r+1} \theta_1 + \dots + \phi_p \theta_{p-r}, \quad r = p-q+1, p-q+2, \dots, p. \end{aligned}$$

Les v_r peuvent être évaluées en pq multiplications.

Chaque élément des $(p+q+1)$ éléments distincts de $F_{\theta\phi}$ est obtenu au moyen de $p+q+1$ multiplications (encore moins si l'on tient compte que $\gamma_r = \gamma_{-r}$) et ainsi le calcul de

tous les éléments de la matrice $F_{\theta\phi}$ nécessite $p^2 + q^2 + 3pq + o_1(p, q)$ opérations et où $o_i(p, q)$ désigne un polynôme de degré i en p et q .

Les $\gamma_h, h = 0, 1, \dots, p + q - 1$, qui sont utilisées dans (4.5.3), (4.5.4) et (4.5.5) peuvent être obtenues efficacement par l'algorithme de Tunncliffe-Wilson (1979) (voir chapitre 1) qui (voir Mélard (1984), pour l'implantation) requiert un nombre d'opérations de $\frac{5}{2}(p + q)^2 + o_1(p, q)$ avec pq multiplications additionnelles pour les coefficients de $\phi(B)\theta(B)$. Le nombre total d'opérations est donc de $4p^2 + 4q^2 + 11pq + o_1(p, q)$. Finalement, on peut donc dire que le nombre total d'opérations est quadratique en p et q .

6. Etude des propriétés statistiques des estimateurs RML et RML_{MZ}

6.1. Généralités

Dans les sections précédentes, nous avons présenté différents algorithmes récurrents. Un problème fondamental qui se pose en lui-même est alors l'étude et l'investigation des *propriétés* de ces algorithmes. Une façon de le faire est d'appliquer l'algorithme en question sur une série de données et d'évaluer par la suite la séquence d'estimateurs obtenus, ce que nous appelons souvent étude par la simulation. Cependant une limitation sérieuse de cette approche est qu'elle est non conclusive, en effet, il sera difficile de déduire à partir des résultats de simulations des conclusions universelles. Pour obtenir des résultats plus généraux nous devons utiliser des analyses, c'est à dire que nous admettons quelques hypothèses sur les données puis nous déterminons, sur la base de celles-ci, les propriétés de cette suite d'estimateurs. Mais comme la relation qui lie les données à la suite des estimateurs est souvent non linéaire et évolutive dans le temps il sera difficile, sauf dans des cas particuliers, d'analyser les propriétés de l'estimateur récurrent. En général, nous ne pouvons décrire analytiquement, que les propriétés asymptotiques de la suite des estimateurs. (i.e. les propriétés lorsque t augmente indéfiniment). Ce paragraphe sera consacré à une description générale des différentes méthodes d'études des propriétés statistiques des algorithmes récurrents étudiés. Nous allons donc énumérer quelques hypothèses pour lesquelles les algorithmes RML et RML_{MZ} convergent, pour une lecture plus détaillée, le lecteur est renvoyé à Ljung et Söderström (1983), Zahaf (1998) et Duflo (1990). Globalement nous pouvons considérer trois principales approches pour analyser l'estimateur récurrent.

1. La première approche consiste à associer une équation différentielle déterministe à une forme générale de l'algorithme d'estimation récurrente donné par Ljung (1977) :

$$\begin{cases} h_t = A(\hat{\beta}_{t-1})h_{t-1} + B(\hat{\beta}_{t-1})\hat{\varepsilon}_t, \\ \hat{\beta}_t = \hat{\beta}_{t-1} + \gamma_t Q(t, \hat{\beta}_{t-1}, h_t), \end{cases}$$

où la fonction h_t est un cas particulier de la quantité d'information S_t , (voir expression (4.2.1)) $Q(\cdot)$ est le vecteur de mise à jour de l'estimateur et les matrices A et B sont des fonctions des paramètres du modèle. La stabilité de cette équation différentielle implique la convergence pour l'algorithme. Cette approche a été suggérée par Ljung (1977) et a été utilisée par de nombreux auteurs, citons en particulier, Söderström et *al.* (1978) et Benveniste et *al.* (1990).

2. Une deuxième approche consiste à introduire une fonction qui joue le rôle de la fonction de Lyapounov (Solo, (1979)) du problème en question. La théorie des martingales permet de déduire la convergence de cette fonction et de là, celle de l'algorithme. Cette approche a été utilisée par Solo (1979) et (1980) et Duflo (1990).

3. La troisième approche qui consiste à étudier l'expression de $R_t \hat{\beta}_t$ n'a pas été appliquée qu'à des algorithmes de types Gauss-Newton dans des cas spéciaux simples, voir Hannan (1980).

6.2. Propriétés statistiques de l'algorithme RML

Pour analyser les propriétés statistiques de l'algorithme RML, Ljung (1977) a utilisé la première approche, en effet il a décrit différents ensembles de conditions possibles (**A**, **B** et **C** voir Zahaf (1998)) pour lesquels l'estimateur $\hat{\beta}_t$ est fortement consistant (théorème 1,2 et 3 de Ljung, 1977). Sous ces hypothèses de régularité, l'estimateur RML est convergent (voir Ljung et Söderström (1983)). La propriété asymptotique est obtenue sous des conditions beaucoup plus restrictives, notamment le fait que la variable aléatoire du processus doit être bornée, ce qui exclut a priori les variables gaussiennes.

6.3. Propriétés statistiques de l'algorithme RML_{MZ}

6.3.1. Préliminaires

En s'inspirant des travaux de Duflo, (1990), Zahaf, (1998) a étudié les propriétés statistiques de l'algorithme RML_{MZ}. En effet, il a exploité la représentation autorégressive du modèle ARMA sous l'hypothèse de l'inversibilité du processus et ce, pour donner une expression de l'erreur de prédiction en fonction des observations. Il a déterminé ensuite une relation entre l'estimateur de Newton, dont les propriétés ont été étudiées par Duflo, (1990) et l'estimateur RML_{MZ} et a montré, en se basant sur la représentation autorégressive du modèle ARMA, que les deux estimateurs ont les mêmes propriétés. Rappelons aussi qu'une méthode récurrente est un algorithme stochastique d'optimisation qui diffère des algorithmes d'optimisation des méthodes hors-ligne par le fait que celle-ci, passe d'un optimum relatif à la série y^t à un optimum relatif à la série y^{t+1} . Or les méthodes stochastiques d'optimisations peuvent être groupées en deux classes : les méthodes du gradient et les méthodes de Newton. (voir chapitre1). Sous quelques hypothèses que nous allons énoncer, Zahaf (1998) montre que l'estimateur RML_{MZ} est égal à l'estimateur de Newton plus une quantité qui tend p. s. vers 0. Ensuite il montre que l'estimateur de Newton est convergent au sens de la convergence p.s. vers β^* .

6.3.2. Convergence de l'estimateur RML_{MZ}

Zahaf, (1998) a considéré les cinq hypothèses suivantes

H1 : Les racines du polynôme autorégressif sont toutes distinctes des racines du polynôme moyenne mobile.

H2 : Les variables aléatoires $\hat{\varepsilon}_t(\beta)$ sont à support borné.

H3 : On suppose l'existence d'un voisinage D_s de β^* tel que pour tout β dans D_s , si l'on considère la représentation AR d'ordre infini de vecteur des paramètres π , alors il existe $o^* > 0$ et il existe $P = P(\beta)$, tel que $\forall i > P, |\pi_i| < o^*$. Nous choisissons alors $P^* = \sup_{\beta \in D_s} P(\beta)$.

H4 : On suppose que $\{\hat{\varepsilon}_t(\beta^*)\}$ est une différence de martingale et de plus

$$E[\varepsilon_{t+1}^2(\beta) | \mathfrak{F}_t] \leq \sigma^2, \text{ pour tout } \beta \text{ dans } D_s$$

H5 : Nous supposons qu'il existe $t_0 > 0$, tel que $\hat{\beta}_t^{NE}$ et $\hat{\beta}_t^{ME}$ se trouvent dans $V \subset D_s$ le

voisinage définit par $v = \left\{ \begin{array}{l} \beta \in R^{p+q} \text{ tel que pour tout } i = 1, \dots, p \text{ et } j = 1, \dots, q, \\ \text{il existe } \delta_{ij} > 0 \text{ et } |a_i(\beta) - b_j(\beta)| > \delta_{ij}, \end{array} \right\}$ où

$a_i(\beta)$ et $b_j(\beta)$ sont les racines des polynômes autorégressif et moyenne mobile respectivement. De plus, pour tout $t > t_0$ si $\hat{\beta}_t^{NE}$ et $\hat{\beta}_t^{ME}$ se trouvent dans $V \subset D_s$, nous appliquons une procédure de projection des deux estimateurs dans cet ensemble.

Les hypothèses **H1** et **H3** garantissent l'inversibilité du processus et donc l'existence d'une représentation autorégressive du processus ARMA. Cependant l'hypothèse **H2** est très classique et utilisée par la plupart des algorithmes récurrents. Par contre les hypothèses **H4** et **H5** sont utilisées pour assurer que les estimateurs RML_{MZ} et de Newton ont les mêmes propriétés.

Proposition

Sous les hypothèses H1, H2, H3, H4 et H5 l'algorithme RML_{MZ} converge presque sûrement vers β^* .

Preuve Voir Zahaf, (1998).

Zahaf, (1998) a montrer aussi la convergence en loi de l'algorithme RML_{MZ} en se basant sur la démonstration du théorème de Robbins-Monro appliqué aux estimateurs de Newton, mais nous n'exposerons pas cette démonstration assez technique, pour une lecture complète nous renvoyons le lecteur vers Zahaf, (1998) et Duflo, (1990).

Chapitre 5

Estimation en-ligne dans les modèles ARMA périodiques

1. Introduction

Nous avons vu au chapitre précédent que l'algorithme récurrent RML appliqué à un modèle ARMA permet de donner des résultats satisfaisants qu'en terme de précision ou de vitesse de convergence (voir aussi Ljung et Soderström, (1983)). Et que son homologue RML_{MZ} , avec des restrictions en plus, permet souvent de donner des résultats encore meilleurs (voir Zahaf, (1998)). Nous allons essayer de prolonger ces algorithmes à une classe particulière de modèles de séries chronologiques à coefficients évolutifs dans le temps à savoir la classe des modèles ARMA périodiques (PARMA).

L'intérêt du présent chapitre sera donc porté sur l'estimation, en présence des données en ligne, d'un modèle autorégressif moyenne mobile périodique (PARMA). Nous tenterons, dans un premier temps, de faire une extension, au modèles PARMA périodiques, des algorithmes : RLS et RML. Nous donnerons ainsi une formulation générale. L'application de l'algorithme RML_{MZ} aux modèles ARMA stationnaires, ayant donnée, des résultats meilleurs que ceux obtenus par l'algorithme RML, nous laisse à croire que la supériorité de RML_{MZ} reste valable pour le cas des modèles PARMA périodiques. Dans cette optique, nous essayerons de généraliser l'algorithme RML_{MZ} pour cette classe de modèles. Pour cela nous aurons besoin, à chaque itération, d'obtenir la matrice d'information de Fisher pour un modèle PARMA, un algorithme pour l'obtention de celle-ci sera introduit en généralisant l'algorithme de Klein et Mélard (1989). Notons aussi que cet algorithme fait intervenir le calcul des autocovariances d'un certain modèle autorégressif périodique pur, à cette fin, nous serons ramenés à construire une procédure généralisant l'algorithme de McLeod (1975) au cas de modèles autorégressifs périodiques. Le plan de ce chapitre est alors le suivant : après cette brève introduction nous présenterons dans une deuxième section les différents principes et critères pris en compte dans la suite de notre travail, nous exposerons ensuite le calcul récursif de l'erreur de prévision dans un modèle PARMA qui nous sera utile ultérieurement. Dans la troisième section sera exposé l'algorithme des moindres carrés récursif, une forme adaptée aux modèles PARMA sera présentée. Le paragraphe qui suit s'intéresse à la formulation de notre adaptation de l'algorithme RML dans le cas périodique. La cinquième section, concernera l'extension de l'algorithme RML_{MZ} , au cas d'un modèle PARMA, notons que ce dernier fait appel au calcul de la matrice d'information de Fisher, d'un modèle PARMA. Un algorithme, basé sur l'extension de l'algorithme de Klein et Mélard, (1989) dans le cas de modèle PARMA, sera présenté, nous terminerons cette section par l'exposé d'une procédure de calcul des autocovariances d'un modèle autorégressif périodique. Enfin, Nous terminons ce chapitre par présenter quelques exemples d'applications.

2. Méthode RPEM dans les modèles ARMA périodiques

2.1. Critères

Nous supposons dans la suite que nous disposons d'une série d'observations y_1, y_2, \dots, y_{Nd} considérée comme réalisation d'un modèle PARMA (Cf. chapitre 2). Notons que dans ce cas les innovations $\{\varepsilon_t, t \in Z\}$ sont de variance périodique dans le temps, le critère défini par (4.2.3) ne sera pas donc adéquat. Une bonne façon pour y remédier sera de faire remplacer le critère (4.2.3) par le critère suivant

$$V_{Nd}(\beta) = \frac{1}{Nd} \sum_{i=1}^{Nd} \frac{1}{\sigma_i^2} \hat{\varepsilon}_i^2(\beta) \quad (5.2.1)$$

où σ_i^2 , la variance de ε_i , est introduite dans le critère (5.2.1) (contrairement au cas AR classique) afin de pondérer les observations. Aström et Bohlin (1965) ont suggéré le critère $1/2 \sum_{i=1}^t \hat{\varepsilon}_i^2(\beta)$ pour l'estimation récursive des paramètres dans les modèles ARMA classiques.

Notons que cette quantité est la même (à une constante près) que celle du critère du maximum de vraisemblance conditionnelle. Les propriétés de l'estimateur obtenu seront équivalentes à celles du maximum de vraisemblance conditionnelle. Ainsi, pour les modèles PARMA nous

avons choisit d'introduire les coefficients $1/\sigma_i^2$ pour préserver la propriété précédente (la même fonction critère que celle du maximum de vraisemblance conditionnelle) ce qui donne le critère suivant : A l'instant $t = j_0 + dm$, l'estimateur $\hat{\beta}_t$ de β , doit minimiser la fonction critère suivante :

$$V_t(\beta) = \frac{1}{2} \sum_{j=1}^t \frac{1}{\sigma_j^2} \hat{\varepsilon}_j^2(\beta) = \begin{cases} \frac{1}{2} \sum_{\tau=0}^{m-1} \sum_{i=1}^d \frac{1}{\sigma_i^2} \hat{\varepsilon}_{i+d\tau}^2 ; & \text{si } j_0 = 0 \\ \frac{1}{2} \sum_{\tau=0}^{m-1} \sum_{i=1}^d \frac{1}{\sigma_i^2} \hat{\varepsilon}_{i+d\tau}^2 + \frac{1}{2} \sum_{j=1}^{j_0} \frac{1}{\sigma_j^2} \hat{\varepsilon}_{j+dm}^2 ; & \text{si } 0 \leq j_0 \leq d-1 \end{cases} \quad (5.2.2)$$

où $\hat{\varepsilon}_t(\beta)$ est l'erreur de prévision.

Nous aurons maintenant besoin d'évaluer l'erreur de prévision avant de passer au schéma général des algorithmes que nous allons présenter, d'où

2.2. Calcul récursif de l'erreur de prévision

Considérons le modèle PARMA $_d(p,q)$, suivant :

$$y_{i+d\tau} - \phi_{i1} y_{i+d\tau-1} - \phi_{i2} y_{i+d\tau-2} - \dots - \phi_{ip} y_{i+d\tau-p} = \varepsilon_{i+d\tau} - \theta_{i1} \varepsilon_{i+d\tau-1} - \dots - \theta_{iq} \varepsilon_{i+d\tau-q} \quad i=1,2,\dots,d \text{ et } \tau \in \mathbb{Z}. \quad (5.2.3)$$

où $\{\varepsilon_t \ni \varepsilon_1, \varepsilon_2, \dots\}$ est un processus bruit blanc périodique.

Notons par $\beta^T = (\phi_{11}, \phi_{12}, \dots, \phi_{1p}, \theta_{11}, \theta_{12}, \dots, \theta_{1q}; \phi_{21}, \dots, \phi_{2p}, \theta_{21}, \dots, \theta_{2q}; \dots, \phi_{d1}, \dots, \phi_{dp}, \theta_{d1}, \dots, \theta_{dq})$ le paramètre vectoriel, à estimer. Alors la prévision, $\hat{y}_{i+d\tau/i+d\tau-1}(\beta)$, de $y_{i+d\tau}$ faite à l'instant $i + d\tau - 1$, est défini par

$$\hat{y}_{i+d\tau/i+d\tau-1}(\beta) = y_{i+d\tau} - \hat{\varepsilon}_{i+d\tau}(\beta) \quad (5.2.4)$$

où $\hat{\varepsilon}_{i+d\tau}(\beta)$ est une approximation de l'innovation $\varepsilon_{i+d\tau}$ qui s'obtient par récurrence sur la base de $y_{i+d\tau-k}$, $k = 1, \dots, p$ et $\hat{\varepsilon}_{i+d\tau-l}$, $l = 1, \dots, q$, selon l'équation suivante :

$$\hat{\varepsilon}_{i+d\tau}(\beta) = y_{i+d\tau} - \phi_{i1} y_{i+d\tau-1} - \dots - \phi_{ip} y_{i+d\tau-p} + \theta_{i1} \hat{\varepsilon}_{i+d\tau-1}(\beta) - \dots - \theta_{iq} \hat{\varepsilon}_{i+d\tau-q}(\beta) \quad i = 1, \dots, d, \tau \in \mathbb{Z} \quad (5.2.5)$$

par la suite, la prévision $\hat{y}_{i+d\tau/i+d\tau-1}(\beta)$ peut se mettre sous la relation

$$\hat{y}_{i+d\tau/i+d\tau-1}(\beta) = \phi_{i1} y_{i+d\tau-1} + \dots + \phi_{ip} y_{i+d\tau-p} - \theta_{i1} \hat{\varepsilon}_{i+d\tau-1}(\beta) - \theta_{i2} \hat{\varepsilon}_{i+d\tau-2}(\beta) - \dots - \theta_{iq} \hat{\varepsilon}_{i+d\tau-q}(\beta) \quad \tau \in \mathbb{Z} \text{ et } i = 1, \dots, d. \quad (5.2.6)$$

Il est nécessaire d'avoir des valeurs initiales mais l'effet de celles-ci décroît en exponentielle si les paramètres vérifient les conditions de causalité et d'inversibilité, donc elles peuvent être choisies arbitrairement sans affecter la prévision. Notons qu'au temps $t = i + d\tau$, les observations $y_{i+d\tau-k}$ étant connues et ce, pour $k = 1, \dots, p$.

Pour avoir une formule de récurrence exprimant la prévision $\hat{y}_{i+d\tau/i+d\tau-1}(\beta)$ nous remplaçons dans l'équation (5.2.6), les termes $\hat{\varepsilon}_{i+d\tau}$ par $y_{i+d\tau} - \hat{y}_{i+d\tau/i+d\tau-1}(\beta)$, nous obtenons ainsi

$$\hat{y}_{t/t-1}(\beta) = \phi_{t1} y_{t-1} + \dots + \phi_{tp} y_{t-p} - \theta_{t1} (y_{t-1} - \hat{y}_{t-1/t-2}(\beta)) - \dots - \theta_{tq} (y_{t-q} - \hat{y}_{t-q/t-q-1}(\beta))$$

finaleme

$$\begin{aligned} \hat{y}_{i+d\tau/i+d\tau-1}(\beta) - \theta_{i1}\hat{y}_{i+d\tau-1/i+d\tau-2}(\beta) - \dots - \theta_{iq}\hat{y}_{i+d\tau-q/i+d\tau-q-1}(\beta) = \\ (\phi_{i1} - \theta_{i1})y_{i+d\tau-1} + \dots + (\phi_{ih} - \theta_{ih})y_{i+d\tau-h}. \\ i = 1, \dots, d. \quad h = \max(p, q). \quad \tau \in \mathbb{Z}. \quad (5.2.7) \end{aligned}$$

Nous avons vu comment obtenir de manière récurrente pour des valeurs des paramètres connus à priori, la prévision $\hat{y}_{t/t-1}(\beta)$ et par conséquent l'erreur de prévision.

3. L'algorithme des moindres carrés récursif (RLS)

L'algorithme RLS (Cf. chapitre 4) peut être considéré, de manière équivalente, comme restriction, au cas de modèles autorégressifs, soit de l'algorithme RML ou bien de l'algorithme PLR (pseudo régression linéaire, voir aussi Ljung et Söderström (1983) et Michaut (1991)). Dans le cas de modèles périodiques la littérature des algorithmes récurrents est relativement maigre comparant au cas ARMA classique, notons exceptionnellement le travail de Adams et Goodwin (1995) consistant à généraliser l'algorithme PLR au cas d'un modèle PARMA. Cependant nous avons choisi une autre formulation particulière, équivalente, en adaptant l'algorithme RLS, présenté dans Ljung et Söderström (1983), au cas d'un modèle PAR. Cet algorithme est dérivé de la méthode usuelle des moindres carrés, mais donné sous une forme récursive. Nous allons tenter de l'appliquer sur la classe des modèles autorégressifs périodiques purs (PAR). Considérons alors le modèle autorégressif périodique d'ordre p suivant (PAR_d(p))

$$y_{i+d\tau} - \phi_{i1}y_{i+d\tau-1} - \dots - \phi_{ip}y_{i+d\tau-p} = \varepsilon_{i+d\tau} \quad i = 1, \dots, d \quad (5.3.1)$$

où $\{\varepsilon_t, t \in \mathbb{Z}\}$ est un bruit blanc périodique et $\beta = (\phi_{11}, \dots, \phi_{1p}, \phi_{21}, \dots, \phi_{1p}, \dots, \phi_{d1}, \dots, \phi_{dp})^T$ est le vecteur des paramètres à estimer. A partir de la donnée de y_t pour $t = 1, \dots, Nd$, introduisons le $p \times d$ -vecteur $\varphi_{i+d\tau}^T = (0, 0, \dots, y_{i+d\tau-1}, y_{i+d\tau-2}, \dots, y_{i+d\tau-p}, 0, \dots, 0)$, $i = 1, \dots, d$, $\tau = 0, \dots, N-1$, qui s'écrit également

$$\varphi_{i+d\tau}(j) = \begin{cases} y_{i+d\tau-j} & \text{si } j = (i-1)p + 1; \dots; ip \\ 0 & \text{sinon} \end{cases}. \quad (5.3.2)$$

alors (5.3.1) peut se réécrire comme suit

$$y_t = \beta^T \varphi_t + \varepsilon_t \quad (5.3.3)$$

Ce modèle décrit la variable y_t comme combinaison linéaire des composantes de φ_t plus un bruit blanc. Une bonne prévision de y_t basée sur son passé que l'on note $\hat{y}_{t/t-1}(\beta)$, est donnée par

$$\hat{y}_{t/t-1}(\beta) = \beta^T \varphi_t \quad (5.3.4)$$

Le critère (5.2.1) devient alors

$$V_{Nd}(\beta) = \frac{1}{Nd} \sum_{i=1}^{Nd} \frac{1}{\sigma_t^2} [y_t - \beta^T \varphi_t]^2 \quad (5.3.5)$$

Pour estimer β , la minimisation de la fonction critère qui, à cause de la linéarité de la prévision par rapport à β , est quadratique, peut être obtenue analytiquement. Ce qui donne

$$\hat{\beta}_{Nd} = \left[\sum_{t=1}^{Nd} \frac{1}{\sigma_t^2} \varphi_t \varphi_t^T \right]^{-1} \sum_{t=1}^{Nd} \frac{1}{\sigma_t^2} \varphi_t y_t \quad (5.3.6)$$

$\hat{\beta}_{Nd}$ est alors défini comme étant l'estimateur des moindres carrés de β .

Notons que le résultat (5.3.6) peut être obtenu de façon récursive, ceci se fera par mise à jour permettant le calcul de $\hat{\beta}_{t+1}$ à partir de $\hat{\beta}_t$ et des nouvelles observations à l'instant $t+1$. Tout en respectant la contrainte de départ (4.2.1). A l'instant $t(i, \tau) = i + d\tau$, notons par

$$\bar{R}_{i+d\tau} = \sum_{k=1}^{i+d\tau} \frac{1}{\sigma_k^2} \varphi_k \varphi_k^T$$

Alors à partir de (5.3.6) il vient que

$$\sum_{k=1}^{i+d\tau} \frac{1}{\sigma_k^2} \varphi_k y_k = \bar{R}_{i+d\tau-1} \hat{\beta}_{i+d\tau-1}$$

à partir de la définition de $\bar{R}_{i+d\tau}$ il s'en suit que

$$\bar{R}_{i+d\tau-1} = \bar{R}_{i+d\tau} - \frac{1}{\sigma_i^2} \varphi_{i+d\tau} \varphi_{i+d\tau}^T \quad (\text{car } \sigma_i^2 \text{ est } d\text{-périodique})$$

donc

$$\begin{aligned} \hat{\beta}_{i+d\tau} &= \bar{R}_{i+d\tau}^{-1} \left[\sum_{k=1}^{i+d\tau-1} \frac{1}{\sigma_k^2} \varphi_k \varphi_k^T + \frac{1}{\sigma_i^2} \varphi_{i+d\tau} \varphi_{i+d\tau}^T \right] \\ &= \bar{R}_{i+d\tau}^{-1} \left[\bar{R}_{i+d\tau-1} \hat{\beta}_{i+d\tau-1} + \frac{1}{\sigma_{i+d\tau}^2} \varphi_{i+d\tau} y_{i+d\tau} \right] \\ &= \bar{R}_{i+d\tau}^{-1} \left[\bar{R}_{i+d\tau} \hat{\beta}_{i+d\tau-1} + \frac{1}{\sigma_i^2} \varphi_{i+d\tau} (-\varphi_{i+d\tau}^T \hat{\beta}_{i+d\tau-1} + y_{i+d\tau}) \right] \\ &= \hat{\beta}_{i+d\tau-1} + \bar{R}_{i+d\tau}^{-1} \varphi_i \frac{1}{\sigma_i^2} [y_{i+d\tau} - \varphi_{i+d\tau}^T \hat{\beta}_{i+d\tau-1}] \end{aligned} \quad (5.3.7)$$

et

$$\bar{R}_{i+d\tau} = \bar{R}_{i+d\tau-1} + \frac{1}{\sigma_i^2} \varphi_{i+d\tau} \varphi_{i+d\tau}^T \quad (5.3.8)$$

Souvent, il est préférable de travailler avec

$$R_{i+d\tau} = \frac{1}{i+d\tau} \bar{R}_{i+d\tau}$$

à partir de (5.3.8) nous trouvons facilement

$$R_{i+d\tau} = \frac{1}{i+d\tau} \left[\bar{R}_{i+d\tau-1} + \frac{1}{\sigma_i^2} \varphi_{i+d\tau} \varphi_{i+d\tau}^T \right] = \frac{i+d\tau-1}{i+d\tau} R_{i+d\tau-1} + \frac{1}{i+d\tau} \frac{1}{\sigma_i^2} \varphi_{i+d\tau} \varphi_{i+d\tau}^T$$

ainsi

$$R_{i+d\tau} = R_{i+d\tau-1} + \frac{1}{i+d\tau} \left[\frac{1}{\sigma_i^2} \varphi_{i+d\tau} \varphi_{i+d\tau}^T - R_{i+d\tau-1} \right] \quad (5.3.9)$$

Les expressions précédentes se résument comme suit

$$\begin{cases} \hat{\beta}_{i+d\tau} = \hat{\beta}_{i+d\tau-1} + \frac{1}{i+d\tau} \frac{1}{\sigma_i^2} R_{i+d\tau}^{-1} \varphi_{i+d\tau} [y_{i+d\tau} - \hat{\beta}_{i+d\tau-1}^T \varphi_{i+d\tau}] \\ R_{i+d\tau} = R_{i+d\tau-1} + \frac{1}{i+d\tau} \left[\frac{1}{\sigma_i^2} \varphi_{i+d\tau} \varphi_{i+d\tau}^T - R_{i+d\tau-1} \right] \end{cases} \quad (5.3.10)$$

Les équations (5.3.10) possèdent la forme souhaitée d'un algorithme récurrent. A l'instant t seulement les valeurs de $\hat{\beta}_t$, R_t , y_t et φ_t vont être stockées en mémoire.

3.1. Estimation récurrente de la variance de l'innovation σ_i^2 $i = 1, \dots, d$

Notons que dans les équations (5.3.10) il est nécessaire d'estimer les paramètres σ_i^2 , $i = 1, \dots, d$. A l'instant $t(i, \tau) = i + d\tau$, notons par $\hat{\sigma}_i^{2(\tau)}$ l'estimateur empirique de σ_i^2 défini par

$$\begin{aligned} \hat{\sigma}_i^{2(\tau)} &= \frac{1}{\tau+1} \sum_{k=0}^{\tau} \hat{\varepsilon}_{i+dk}^2 \\ &= \frac{1}{\tau+1} \left[\sum_{k=0}^{\tau-1} \hat{\varepsilon}_{i+dk}^2 + \hat{\varepsilon}_{i+d\tau}^2 \right] \end{aligned}$$

d'où

$$\hat{\sigma}_i^{2(\tau)} = \frac{\tau}{\tau+1} \hat{\sigma}_i^{2(\tau-1)} + \frac{1}{\tau+1} \hat{\varepsilon}_{i+d\tau}^2 \quad (5.3.11)$$

Cependant l'algorithme décrit par les équations (5.3.10) peut être amélioré du point de vue numérique en introduisant

$$P_t = \bar{R}_t^{-1} = \frac{1}{t} R_t^{-1} \quad (5.3.12)$$

Puis de calculer P_t directement à partir des équations de récurrence (5.3.10). Cela peut être accompli au moyen du résultat connu sous le nom du lemme d'inversion matricielle, que nous avons déjà présenté dans le chapitre 4, et que nous l'énonçons sous une autre variante

Lemme 5.3.1

Soient A, B, C et D des matrices de dimensions compatibles de sorte que la somme BCD et la somme $A + BCD$ existent. Alors

$$[A + BCD]^{-1} = A^{-1} - A^{-1}B[DA^{-1}B + C^{-1}]^{-1}DA^{-1} \quad (5.3.13)$$

Appliquons (5.3.13) à (5.3.7) et (5.3.8) avec

$$A = P_{i+d\tau}, \quad B = \varphi_{i+d\tau}, \quad C = \frac{1}{\sigma_i^2}, \quad D = \varphi_{i+d\tau}^T$$

Il s'en suit alors que

$$P_{i+d\tau} = \left[P_{i+d\tau-1}^{-1} + \varphi_{i+d\tau} \frac{1}{\sigma_i^2} \varphi_{i+d\tau}^T \right]^{-1}$$

$$= P_{i+d\tau-1} - \frac{P_{i+d\tau-1}\varphi_{i+d\tau}\varphi_{i+d\tau}^T P_{i+d\tau-1}}{\sigma_{i+d\tau}^2 + \varphi_{i+d\tau}^T P_{i+d\tau-1}\varphi_{i+d\tau}} \quad (5.3.14)$$

Comme dans le chapitre 4, l'avantage de (5.3.14) para rapport à (5.3.9), est la réduction de la dimension de la matrice à inverser (inversion d'un scalaire au lieu de l'inversion d'une matrice d'ordre $\dim \beta$).

$$\begin{aligned} \frac{1}{\sigma_i^2} P_{i+d\tau}\varphi_{i+d\tau} &= \frac{1}{\sigma_i^2} P_{i+d\tau-1}\varphi_{i+d\tau} - \frac{P_{i+d\tau-1}\varphi_{i+d\tau}\varphi_{i+d\tau}^T P_{i+d\tau-1}\varphi_{i+d\tau}}{\sigma_i^2 + \varphi_{i+d\tau}^T P_{i+d\tau-1}\varphi_{i+d\tau}} \\ &= \frac{P_{i+d\tau-1}\varphi_{i+d\tau}}{\sigma_i^2 + \varphi_{i+d\tau}^T P_{i+d\tau-1}\varphi_{i+d\tau}} \end{aligned} \quad (5.3.15)$$

Ainsi nous obtenons la forme complète d'implémentation numérique de l'algorithme ci-dessous.

$$\begin{cases} \hat{\varepsilon}_{i+d\tau} = y_{i+d\tau} - \hat{\beta}_{i+d\tau-1}\varphi_{i+d\tau} \\ \hat{\sigma}_i^{2(\tau)} = \frac{\tau}{\tau+1}\hat{\sigma}_i^{2(\tau-1)} + \frac{1}{\tau+1}\hat{\varepsilon}_{i+d\tau}^2 \\ L_{i+d\tau} = \frac{P_{i+d\tau-1}\varphi_{i+d\tau}}{\hat{\sigma}_i^{2(\tau)} + \varphi_{i+d\tau}^T P_{i+d\tau-1}\varphi_{i+d\tau}} \\ \hat{\beta}_{i+d\tau} = \hat{\beta}_{i+d\tau-1} + L_{i+d\tau}\varepsilon_{i+d\tau} \\ P_{i+d\tau} = P_{i+d\tau-1} - L_{i+d\tau}\varphi_{i+d\tau}^T P_{i+d\tau-1} \end{cases} \quad (5.3.16)$$

où φ_i est donné par (5.3.2).

Les équations (5.3.16) définissent l'algorithme des moindres carrés récursif (RLS) adapté aux modèles PARMA. Celui-ci peut être l'un des algorithmes les plus utilisés dans les applications industrielles en raison de son extrême simplicité de mise en œuvre, de sa robustesse aux erreurs de calcul et de la diversité des contextes de son utilisation.

Les opérations de calcul intervenant dans les équations (5.3.16) nécessitent dans ce cas $(p \times d)^2$ multiplications, et mémorisation de vecteurs de taille $p \times d$ et d'une matrice de taille p^2 . La complexité temporelle et spatiale sera donc de l'ordre de $O((p \times d)^2)$.

4. Extension de l'algorithme RML au cas périodique

Dans ce paragraphe nous allons adapter l'algorithme RML, déjà présenté dans le cas ARMA classique au chapitre 4, aux modèles ARMA périodiques. En gardant le même principe, nous choisissons alors le critère (5.2.2) c'est à dire celui minimisant la somme pondérée des carrés des erreurs de prévision.

Soit $\hat{\beta}_{t-1}$ l'estimateur de β à l'instant $t-1$, c'est à dire celui qui minimise la quantité $V_{t-1}(\beta)$. Nous voulons obtenir, à l'instant t , l'estimateur $\hat{\beta}_t$ qui (approximativement) minimise la fonction critère. Nous devons assurer deux hypothèses fondamentales sous lesquelles l'algorithme que nous allons présenter donne un estimateur optimal.

B1 : Nous supposons que $\hat{\beta}_t$ se trouve dans un voisinage de $\hat{\beta}_{t-1}$. Et que les dérivées secondes sont continues.

B2 : Nous supposons également que $\hat{\beta}_{t-1}$ est bien l'estimateur optimal à l'instant $t-1$.

Pour $t = i + d\tau$, notons par : $\psi_t(\beta)$ l'opposé de la dérivée de $\hat{\varepsilon}_t(\beta)$ par rapport à β

$$\psi_t(\beta) = \left[-\frac{\partial \hat{\varepsilon}_t(\beta)}{\partial \beta^T} \right]^T, \quad \bar{R}_t \text{ une approximation de la matrice des dérivées secondes } \frac{\partial^2 V_t(\beta_t)}{\partial \beta \partial \beta^T}$$

et introduisons le $(p+q) \times d$ -vecteur $\varphi_{i+d\tau}$, relatif à la période i définit par :

$$\varphi_{i+d\tau}(j) = \begin{cases} y_{i+d\tau-j} & \text{si } j = (i-1)q + (i-1)p + 1; \dots; (i-1)q + ip \\ -\varepsilon_{i+d\tau-j} & \text{si } j = (i-1)q + ip + 1; \dots; iq + ip \\ 0 & \text{sinon} \end{cases}. \quad (5.4.0)$$

$i = 1, \dots, d$, alors

Proposition 5.4.1

Sous les Hypothèses **B1** et **B2** l'algorithme décrit par les équations qui suivent donne un estimateur dont la fonction critère (5.2.2) est optimale.

$$\left\{ \begin{array}{l} \hat{\varepsilon}_{i+d\tau} = y_{i+d\tau} - \hat{\beta}_{d\tau}^T \varphi_{i+d\tau}, \\ \hat{\sigma}_i^{2(\tau)} = \frac{\tau}{\tau+1} \hat{\sigma}_i^{2(\tau-1)} + \frac{1}{\tau+1} \hat{\varepsilon}_{i+d\tau}^2 \\ \hat{\beta}_{i+d\tau} = \hat{\beta}_{i+d\tau-1} + \frac{1}{\hat{\sigma}_i^{(\tau)}} \bar{R}_{i+d\tau}^{-1} \psi_{i+d\tau} \hat{\varepsilon}_{i+d\tau}, \\ \psi_{i+1+d\tau} = \hat{\theta}_{i1}^{t-1} \psi_{i+d\tau} + \hat{\theta}_{i2}^{t-1} \psi_{i-1+d\tau} + \dots + \hat{\theta}_{iq}^{t-1} \psi_{i+d\tau-q+1}, \\ \bar{R}_{i+1+d\tau} = \bar{R}_{i+d\tau} + \frac{1}{\hat{\sigma}_i^{(\tau+1)}} \psi_{i+1+d\tau} \psi_{i+1+d\tau}^T. \end{array} \right.$$

pour $i = 1, \dots, d$, $\tau \in \mathbb{Z}$.

Pour une étude détaillée des propriétés statistiques (convergence vers la vraie valeur du paramètre β , etc.) de l'estimateur obtenu de cet algorithme, le lecteur sera renvoyé à Ljung et Soderström (1983), Duflo (1990) ou Adams et Goodwin, (1995).

preuve

Développons par la formule de Taylor à l'ordre 2 la fonction $V_t(\beta)$ autour de $\hat{\beta}_{t-1}$ nous aurons

$$V_t(\beta) = V_t(\hat{\beta}_{t-1}) + \frac{\partial V_t(\hat{\beta}_{t-1})}{\partial \beta^T} [\beta - \hat{\beta}_{t-1}] + \frac{1}{2} [\beta - \hat{\beta}_{t-1}]^T \frac{\partial^2 V_t(\hat{\beta}_{t-1})}{\partial \beta \partial \beta^T} [\beta - \hat{\beta}_{t-1}] + o(|\beta - \hat{\beta}_{t-1}|^2), \quad (5.4.1)$$

où $(\partial./\partial \beta^T)$ note la différentiation par rapport à β et $o(x)$ note une fonction telle que $\frac{o(x)}{|x|} \rightarrow 0$ quand $|x| \rightarrow 0$. Pour minimiser la fonction $V_t(\beta)$, nous dérivons le membre droit

de l'équation (5.4.1) par rapport à β , l'expression obtenue est mise égale à zéro. Nous obtenons donc

$$\hat{\beta}_t = \hat{\beta}_{t-1} - \left[\frac{\partial^2 V_t(\hat{\beta}_{t-1})}{\partial \beta \partial \beta^T} \right]^{-1} \left[\frac{\partial V_t(\hat{\beta}_{t-1})}{\partial \beta^T} \right]^T + o(|\hat{\beta}_t - \hat{\beta}_{t-1}|) \quad (5.4.2)$$

Cette dernière équation, semblant déterminer le vecteur des paramètres β de façon récursive, n'est pas tout à fait claire, nous allons faire une série d'approximations pour l'évaluer. Nous avons que

$$\psi_t(\beta) = \left[-\frac{\partial \hat{\varepsilon}_t(\beta)}{\partial \beta^T} \right]^T \quad (5.4.3)$$

à partir de (5.2.1) nous pouvons écrire

$$\left[\frac{\partial V_t(\beta)}{\partial \beta^T} \right]^T = -\sum_{k=1}^t \frac{1}{\sigma_k^2} \psi_k(\beta) \hat{\varepsilon}_k(\beta) = \left[\frac{\partial V_{t-1}(\beta)}{\partial \beta^T} \right]^T - \frac{1}{\sigma_t^2} \psi_t(\beta) \hat{\varepsilon}_t(\beta) \quad (5.4.4)$$

et par différentiation une fois de plus,

$$\frac{\partial^2 V_t(\beta)}{\partial \beta \partial \beta^T} = \frac{\partial^2 V_{t-1}(\beta)}{\partial \beta \partial \beta^T} + \frac{1}{\sigma_t^2} \psi_t(\beta) \psi_t^T(\beta) + \frac{1}{\sigma_t^2} \frac{\partial^2 \hat{\varepsilon}_t(\beta)}{\partial \beta \partial \beta^T} \hat{\varepsilon}_t(\beta) \quad (5.4.5)$$

Pour évaluer (5.4.2), un certain nombre d'approximations est introduit, dans l'ordre.

- La supposition **B1** permet d'écrire les deux approximations suivantes :

$$\text{Négligeons } o(|\hat{\beta}_t - \hat{\beta}_{t-1}|) \text{ dans (5.4.2)} \quad (5.4.6)$$

Et à cause de la continuité des dérivées secondes à l'instant t , prenons

$$\frac{\partial^2 V_t(\hat{\beta}_t)}{\partial \beta \partial \beta^T} = \frac{\partial^2 V_t(\hat{\beta}_{t-1})}{\partial \beta \partial \beta^T} \quad (5.4.7)$$

- En se servant de **B2** nous pourrions écrire tout de suite

$$\frac{\partial V_{t-1}(\hat{\beta}_{t-1})}{\partial \beta^T} = 0 \quad (5.4.8)$$

- Posons enfin

$$\frac{\partial^2 \hat{\varepsilon}_t(\hat{\beta}_{t-1})}{\partial \beta \partial \beta^T} \hat{\varepsilon}_t(\hat{\beta}_{t-1}) \approx 0 \quad (5.4.9)$$

La raison de cette dernière approximation est la suivante. Prés de la valeur β^* , $\{\hat{\varepsilon}_t(\beta)\}$ sera un processus bruit blanc périodique, ainsi nous pouvons approximativement considérer que $\hat{\varepsilon}_t(\beta)$ est de moyenne nulle et indépendant de ce qui c'est passé avant le temps $t-1$. En effet, la dérivée seconde de $\hat{\varepsilon}_t(\beta)$ par rapport à β est égale à l'opposé de celle de $\hat{y}_{t/t-1}(\beta)$, par ailleurs la prévision $\hat{y}_{t/t-1}(\beta)$ s'exprime en combinaison linéaire du passé de y_t , comme $\hat{\varepsilon}_t(\beta)$ est l'innovation du processus les variables aléatoires $\hat{\varepsilon}_t(\beta)$ et $\partial^2 \hat{\varepsilon}_t(\beta) / \partial \beta \partial \beta^T$ seront donc non corrélées. Et par conséquent l'espérance mathématique du membre de gauche de l'expression (5.4.9) est alors bien proche de 0, ainsi le dernier terme de (5.4.5) contribue beaucoup moins à $(\partial^2 V_t(\beta) / \partial \beta \partial \beta^T)$ que le second terme.

Sous les suppositions (5.4.9) et (5.4.7) insérées dans (5.4.5) nous pouvons évaluer approximativement la matrice dérivée seconde. Notons cette matrice par \bar{R}_t , alors nous avons

$$\bar{R}_t = \bar{R}_{t-1} + \frac{1}{\sigma_t^2} \psi_t(\hat{\beta}_{t-1}) \psi_t^T(\hat{\beta}_{t-1}) \quad (5.4.10)$$

En se servant de (5.4.8) et de (5.4.4), nous aurons

$$\left[\frac{\partial V_t(\hat{\beta}_{t-1})}{\partial \beta} \right]^T = -\frac{1}{\sigma_t^2} \psi_t(\hat{\beta}_{t-1}) \hat{\varepsilon}_t(\hat{\beta}_{t-1}) \quad (5.4.11)$$

Utilisons cette dernière équation et l'approximation de (5.4.10) dans (5.4.1), nous aurons le système

$$\begin{cases} \bar{R}_t = \bar{R}_{t-1} + \frac{1}{\sigma_t^2} \psi_t(\hat{\beta}_{t-1}) \psi_t^T(\hat{\beta}_{t-1}) \\ \hat{\beta}_t = \hat{\beta}_{t-1} + \frac{1}{\sigma_t^2} \bar{R}_t^{-1} \psi_t(\hat{\beta}_{t-1}) \hat{\varepsilon}_t(\hat{\beta}_{t-1}) \end{cases} \quad (5.4.12)$$

Ce système n'est pas encore exploitable, il reste maintenant, de discuter dans ce cas comment déterminer $\hat{\varepsilon}_t(\hat{\beta}_{t-1})$ et $\psi_t(\hat{\beta}_{t-1})$.

4.1. Approximation de $\hat{\varepsilon}_t(\hat{\beta}_{t-1})$ et de $\psi_t(\hat{\beta}_{t-1})$

Si nous considérons le modèle PARMA $_d(p,q)$, d -périodique, de (3.4.1) alors le vecteur des paramètres β^T s'identifie par :

$\beta^T = (\phi_{11}, \phi_{12}, \dots, \phi_{1p}, \theta_{11}, \theta_{12}, \dots, \theta_{1q}; \phi_{21}, \dots, \phi_{2p}, \theta_{21}, \dots, \theta_{2q}; \dots; \phi_{d1}, \dots, \phi_{dp}, \theta_{d1}, \dots, \theta_{dq})$,
ainsi nous pouvons écrire

$$\psi_t^T(\beta) = -\frac{\partial \hat{\varepsilon}_t(\beta)}{\partial \beta} = \frac{\partial \hat{y}_{t/t-1}(\beta)}{\partial \beta^T} \quad \text{pour } t = 1 + d\tau, 2 + d\tau, \dots, d + d\tau. \quad (5.4.13)$$

A partir de (5.2.7), et par dérivation par rapport aux paramètres ϕ_{jk} $j = 1, \dots, d$, $k = 1, \dots, p$ et θ_{lm} $l = 1, \dots, d$ $m = 1, \dots, q$ nous trouvons directement

$$\begin{aligned} \frac{\partial \hat{y}_{i+d\tau/i+d\tau-1}(\beta)}{\partial \phi_{jk}} - \theta_{i1} \frac{\partial \hat{y}_{i+d\tau-1/i+d\tau-2}(\beta)}{\partial \phi_{jk}} - \dots - \theta_{iq} \frac{\partial \hat{y}_{i+d\tau-q/i+d\tau-q-1}(\beta)}{\partial \phi_{jk}} &= \begin{cases} y_{i+d\tau-k} & \text{si } j = i \\ 0 & \text{sinon} \end{cases} \\ \frac{\partial \hat{y}_{i+d\tau/i+d\tau-1}(\beta)}{\partial \theta_{lm}} - \theta_{i1} \frac{\partial \hat{y}_{i+d\tau-1/i+d\tau-2}(\beta)}{\partial \theta_{lm}} - \dots - \theta_{iq} \frac{\partial \hat{y}_{i+d\tau-q/i+d\tau-q-1}(\beta)}{\partial \theta_{lm}} &= \begin{cases} -\hat{\varepsilon}_{i+d\tau-m}(\beta) & \text{si } l = i \\ 0 & \text{sinon} \end{cases} \end{aligned} \quad (5.4.14)$$

Nous allons donner des approximations de $\hat{\varepsilon}_t(\hat{\beta}_{t-1})$ et $\psi_t(\hat{\beta}_{t-1})$ qui peuvent être calculées par récurrence au sens de (4.2.1).

Une approximation raisonnable de $\hat{\varepsilon}_t(\hat{\beta}_{t-1})$ sera $\hat{\varepsilon}_t$, calculé suivant

$$\hat{\varepsilon}_t = y_t - \hat{y}_{t/t-1} \quad (5.4.15)$$

où $\hat{y}_{t/t-1}$ est calculée selon la formule (5.2.7) en remplaçant β par $\hat{\beta}_{t-1}$ avec

$$\hat{\beta}_{t-1} = (\hat{\phi}_{11}^{t-1}, \dots, \hat{\phi}_{1p}^{t-1}, \hat{\theta}_{11}^{t-1}, \dots, \hat{\theta}_{1q}^{t-1}; \hat{\phi}_{21}^{t-1}, \dots, \hat{\phi}_{2p}^{t-1}, \hat{\theta}_{21}^{t-1}, \dots, \hat{\theta}_{2q}^{t-1}; \dots; \hat{\phi}_{d1}^{t-1}, \dots, \hat{\phi}_{dp}^{t-1}, \hat{\theta}_{d1}^{t-1}, \dots, \hat{\theta}_{dq}^{t-1})^T$$

c'est à dire que

$$\begin{aligned} \hat{y}_{i+d\tau/i+d\tau-1} &= \hat{\theta}_{i1}^{t-1} \hat{y}_{i+d\tau-1/i+d\tau-2} + \dots + \hat{\theta}_{iq}^{t-1} \hat{y}_{i+d\tau-q/i+d\tau-q-1} + (-\hat{\theta}_{i1}^{t-1} + \hat{\phi}_{i1}^{t-1}) y_{i-1+d\tau} + \dots \\ &\quad \dots + (-\hat{\theta}_{ih}^{t-1} + \hat{\phi}_{ih}^{t-1}) y_{i+d\tau-h} \\ &\quad i = 1, \dots, d. \end{aligned}$$

Alors nous pouvons écrire

$$\hat{y}_{i+d\tau/i+d\tau-1}(\beta) = \hat{\beta}_{i-1+d\tau}^T \varphi_{i+d\tau}$$

et par conséquent

$$\hat{\varepsilon}_{i+d\tau} = y_{i+d\tau} - \hat{\beta}_{i-1+d\tau}^T \varphi_{i+d\tau} \quad i = 1, \dots, d. \quad (5.4.16)$$

De même, nous approximations $\psi_t(\hat{\beta}_{t-1})$ par ψ_t calculée à partir de la formule suivante

$$\psi_{i+d\tau} = \hat{\theta}_{i1}^{t-1} \psi_{i-1+d\tau} + \hat{\theta}_{i2}^{t-1} \psi_{i-2+d\tau} + \dots + \hat{\theta}_{iq}^{t-1} \psi_{i-q+d\tau} + \varphi_{i+d\tau-q} \quad i = 1, \dots, d. \quad (5.4.17)$$

Donc $\hat{\varepsilon}_t$ et ψ_t seront calculés par récurrence à partir de (5.4.16) et (5.4.17). Au temps t nous avons seulement besoin de connaître φ_t, ψ_t et $\hat{\beta}_{t-1}$.

Pour estimer, de façon récurrente, la variance périodique σ_i^2 $i = 1, \dots, d$, du bruit blanc nous nous servons de (5.3.11).

En appliquant le lemme d'inversion matricielle nous retrouvons les équations de l'algorithme RML périodique appliqué à un modèle PARMA $_d(p, q)$

$$\left\{ \begin{array}{l} \hat{\varepsilon}_{i+d\tau} = y_{i+d\tau} - \hat{\beta}_{d\tau}^T \varphi_{i+d\tau}, \\ \hat{\sigma}_i^{2(\tau)} = \frac{\tau}{\tau+1} \hat{\sigma}_i^{2(\tau-1)} + \frac{1}{\tau+1} \hat{\varepsilon}_{i+d\tau}^2, \\ \hat{\beta}_{i+d\tau} = \hat{\beta}_{i+d\tau-1} + \frac{1}{\hat{\sigma}_i^{2(\tau)}} \bar{R}_{i+d\tau}^{-1} \psi_{i+d\tau} \hat{\varepsilon}_{i+d\tau}, \\ \psi_{i+1+d\tau} = \hat{\theta}_{i1}^{t-1} \psi_{i+d\tau} + \hat{\theta}_{i2}^{t-1} \psi_{i-1+d\tau} + \dots + \hat{\theta}_{iq}^{t-1} \psi_{i-q+d\tau} + \varphi_{i+d\tau-q+1}, \\ \bar{R}_{i+1+d\tau} = \bar{R}_{i+d\tau} + \frac{1}{\hat{\sigma}_i^{2(\tau+1)}} \psi_{i+1+d\tau} \psi_{i+1+d\tau}^T. \end{array} \right. \quad (5.4.18)$$

pour $i = 1, \dots, d$ et $\tau \in \mathbb{Z}$, où $\varphi_{i+d\tau}$ est donné par (5.4.0).

5. Extension de l'algorithme RML_{MZ} au cas périodique

Nous allons maintenant étendre l'algorithme RML_{MZ} introduit par Zahaf (1998). Comme nous avons déjà vu au chapitre 4, cet algorithme est basé sur la matrice d'information de Fisher des paramètre du modèle en question. Nous commençons d'abord par construire un algorithme basé sur l'algorithme de Klein et Mélard (1989) pour l'obtention de la matrice d'information de Fisher d'un modèle ARMA périodique.

5.1. Calcul de la matrice d'information de Fisher pour un modèle ARMA périodique

Dans les problèmes d'estimation des modèles de séries chronologiques, l'importance de la matrice d'information de Fisher n'est plus à démontrer. En effet, la littérature est surabondante et son intérêt n'a pas cessé de croître. Pour beaucoup de modèles employés, cette mesure d'information est souvent considérée comme une bonne approximation de la covariance asymptotique de l'estimateur du maximum de vraisemblance gaussien. La connaissance de la structure de celle-ci est très utile pour les méthodes d'estimation qui, dans le cas du maximum de vraisemblance, sont de nature itérative. D'autres part la l'obtention de la matrice de covariance d'un estimateur est très nécessaire dans les problèmes inférentiels, tels que les problèmes de tests sur les paramètres, les problèmes de la détermination de la taille de l'échantillon, problème de scoring, etc. Par ailleurs, l'utilisation de la matrice d'information de Fisher est très répandue dans le domaine du contrôle adaptatif et du traitement de signal, en particulier, pour les algorithmes récurrents de l'estimation des

paramètres qui, à chaque itération, requièrent le calcul de cette matrice (l'algorithme de Newton, Duflo, (1990)).

La matrice d'information de Fisher a un long historique. Il s'agissait tout d'abord de la matrice d'information asymptotique, celle-ci a été introduite pour la première fois par Whittle (1953), qui s'est basée sur une approximation de la fonction de vraisemblance Gaussienne assurant une validation asymptotique. Par la suite plusieurs auteurs ont donné des développements plus importants (pour une revue plus détaillée voir Zahaf (1998)). Cependant la matrice d'information exacte est définie comme l'opposé de l'espérance mathématique de la hessienne de la fonction de vraisemblance, évaluée en la valeur finale des paramètres. Plus récemment plusieurs travaux ont été portés sur la matrice exacte. Porat et Friedlander (1986b) ont décrit, à ce sujet, un algorithme pour un modèle ARMA univarié avec une composante additive déterministe. La méthode est à la fois complexe et très coûteuse en calcul. Le nombre d'opérations scalaires (nombre de multiplications et d'additions) est de l'ordre de N^2 . Indépendamment, Zadrozny (1989, 1992) et Teircerco (1990) ont donné un algorithme bien plus efficace, puisque le nombre d'opérations est proportionnel à N . La méthode est basée sur le *filtre de Kalman*, et a été appliquée au modèle VARMA par le premier, et à la forme *d'espace d'état* par le second. Mélard et Klein (1994) ont décrit une méthode pour calculer la matrice d'information d'un modèle ARMA univarié. Cette méthode est basée sur l'expression alternative de la fonction exacte de vraisemblance gaussienne au moyen des équations de Chandrasekhar au lieu des équations du filtre de Kalman. Toujours dans le cas univarié, Klein, Mélard et Zahaf (1998) ont amélioré et généralisé la méthode pour SISO (single input single output). Ces deux algorithmes nécessitent un nombre d'opérations proportionnel à N . Zahaf (1998) a proposé une généralisation de la méthode de calcul de cette matrice pour le modèle SISO au modèle multivarié général d'espace d'état.

Dans la littérature relative aux modèles PARMA périodiques, se trouve un résultat donnant l'expression exacte de la matrice d'information de Fisher (voir Pagano (1978)) concernant seulement le cas des modèles autorégressifs périodiques purs (PAR_d). De plus ce résultat ne donne pas les détails techniques de l'obtention de cette matrice mais des résultats généraux. Cependant il n'existe pas d'algorithmes précis traitant le cas d'un modèle PARMA périodique, Sauf si celui-ci est transformé en sa forme ARMA multivariée équivalente. Il est possible, à ce moment là, d'appliquer l'un des algorithmes existants. Il faut souligner qu'en procédant de la sorte, le coût de calcul sera nettement supérieur au coût relatif au cas où l'on généraliserait, aux modèles périodiques, une des méthodes classiques. Nous avons donc choisi d'étendre l'algorithme de Klein et Mélard (1989) qui calcul la matrice d'information relative aux paramètres autorégressifs et moyenne mobile seulement, d'un modèle ARMA univarié. Juste avant, nous allons énoncer le résultat de Pagano (1978) pour les processus autorégressifs périodiques.

5.1.1. Calcul de la matrice d'information de Fisher d'un modèle PAR_d(p)

Considérons le modèle PAR_d(p) de (5.3.1).

Notons par $\Phi = (\phi_{11}, \dots, \phi_{1p}, \phi_{21}, \dots, \phi_{2p}, \dots, \phi_{d1}, \dots, \phi_{dp})^T$ et $\sigma = (\sigma_1^2, \sigma_2^2, \dots, \sigma_d^2)^T$ les vecteurs des paramètres du modèle. Par définition, la matrice d'information de Fisher pour les paramètres Φ et relative à une série, y^{Nd} , de taille Nd d'un modèle PAR_d(p) est donnée par :

$$F_{Nd} = -E \left[\frac{\partial^2 \log l(\Phi)}{\partial \Phi \partial \Phi^T} \right], \quad (5.5.1)$$

où $l(\Phi)$ est la fonction de vraisemblance Gaussienne de Φ relative à une série y_1, y_2, \dots, y_{Nd} qui est supposée de taille multiple de la période d . Notons par F , la matrice d'information de Fisher limite (voir Whittle, (1953)), alors d'après Pagano, (1978) nous avons $F = (1/N) F_{Nd}$

Nous pouvons considérer la partition suivante

$$F = \begin{pmatrix} F_{11} & F_{12} & \cdots & F_{1d} \\ F_{21} & F_{22} & \cdots & F_{2d} \\ \vdots & \vdots & \vdots & \vdots \\ F_{d1} & F_{d2} & \cdots & F_{dd} \end{pmatrix} = (F_{jl})_{j,l=1,\dots,d} \quad (5.5.2)$$

$$\text{où } (F_{jl})_{km} = F(\phi_{jk}, \phi_{lm}) = -E \left(\frac{\partial^2 \log l(\Phi)}{\partial \phi_{jk} \partial \phi_{lm}} \right) \quad k = 1, \dots, p, \quad m = 1, \dots, p$$

D'après Pagano (1978), l'estimateur des moments $\tilde{\Phi}$ de Φ se distribue asymptotiquement, à une constante près, selon une loi normale de moyenne nulle et de matrice covariances, l'inverse de la matrice d'information de Fisher, l'expression de cette matrice est :

$$F(\phi_{kj}, \phi_{ml}) = \delta_{km} \gamma(k-j, k-l) / \sigma_k^2 \quad (5.5.3)$$

pour $k, m = 1, \dots, d$ et $j, l = 1, \dots, p$, où δ est le symbole de Kronicker, $\delta_{km} = \begin{cases} 1 & \text{si } k = m \\ 0 & \text{sinon} \end{cases}$.

D'après (2.4), la matrice F peut se partitionner comme suit

$$F = \begin{pmatrix} F_{11} & 0 & 0 & 0 \\ 0 & F_{22} & 0 & 0 \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & F_{dd} \end{pmatrix} \quad \text{d'où} \quad F^{-1} = \begin{pmatrix} F_{11}^{-1} & 0 & 0 & 0 \\ 0 & F_{22}^{-1} & 0 & 0 \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & F_{dd}^{-1} \end{pmatrix} \quad (5.5.4)$$

Notons en premier lieu que la matrice F est symétrique et diagonale par bloc, donc son inverse est facilement calculable relativement à une matrice quelconque.

Pour déterminer les éléments de la matrice nous devons déterminer une estimation des autocovariances $\gamma^{(\cdot)}$. Celle-ci s'obtient en fonction de l'instance du problème traité. En effet, soit directement sur la base des observations, dans ce cas nous utiliserons les autocovariances empiriques (voir Vecchia 1985b). Soit sur la base des paramètres Φ et σ du modèle qui seront supposés connus. Dans ce dernier cas les estimations seront obtenues au moyen des équations de Yule-Walker périodiques (Voir section 5.4).

5.1.2 Calcul de la matrice d'information de Fisher d'un modèle PARMA $_d(p,q)$ général

Lorsque nous rajoutons une composante moyenne mobile, en général nous n'avons pas de formules simples analogues à (5.1.3) pour l'obtention de la matrice F . Nous pouvons cependant, adapter l'un des algorithmes disponibles pour le calcul des éléments de celle-ci dans le cas d'un modèle ARMA classique, nous avons choisi l'algorithme de Klein et Mélard (1989).

Considérons le modèle PARMA $_d(p,q)$ de (5.2.13). Notons Maintenant par $\sigma = (\sigma_1^2, \sigma_2^2, \dots, \sigma_d^2)^T$ et

$$\beta^T = (\phi_{11}, \phi_{12}, \dots, \phi_{1p}, \theta_{11}, \theta_{12}, \dots, \theta_{1q}; \phi_{21}, \dots, \phi_{2p}, \theta_{21}, \dots, \theta_{2q}; \dots; \phi_{d1}, \dots, \phi_{dp}, \theta_{d1}, \dots, \theta_{dq}) \text{ les vecteurs}$$

des paramètres du modèle. La matrice d'information de Fisher pour les paramètres autorégressifs et moyenne mobile d'un modèle ARMA $_d(p,q)$ est donnée par (5.5.1) (en remplaçant Φ par β). Comme dans le paragraphe précédent, la matrice F peut se mettre sous la forme suivante

$$F = \begin{pmatrix} F_{11} & F_{12} & \cdots & F_{1d} \\ F_{21} & F_{22} & \cdots & F_{2d} \\ \vdots & \vdots & \vdots & \vdots \\ F_{d1} & F_{d2} & \cdots & F_{dd} \end{pmatrix} = (F_{jl})_{j,l=1,\dots,d} \quad (5.5.5)$$

où $F_{jl} = \begin{pmatrix} F_{\Phi_j\Phi_l} & F_{\Phi_j\Theta_l} \\ F_{\Theta_j\Phi_l} & F_{\Theta_j\Theta_l} \end{pmatrix}$

les matrices $F_{\Phi_j\Phi_l}$, $F_{\Phi_j\Theta_l}$, $F_{\Theta_j\Phi_l}$, $F_{\Theta_j\Theta_l}$ sont définies par

$$\begin{aligned} (F_{\Phi_j\Phi_l})_{km} &= F(\phi_{jk}, \phi_{lm}) = -E \left(\frac{\partial^2 \log l(\beta)}{\partial \phi_{jk} \partial \phi_{lm}} \right) \quad k=1,\dots,p, m=1,\dots,p \\ (F_{\Phi_j\Theta_l})_{km} &= F(\phi_{jk}, \theta_{lm}) = -E \left(\frac{\partial^2 \log l(\beta)}{\partial \phi_{jk} \partial \theta_{lm}} \right) \quad k=1,\dots,p, m=1,\dots,q \\ (F_{\Theta_j\Phi_l})_{km} &= F(\theta_{jk}, \phi_{lm}) = -E \left(\frac{\partial^2 \log l(\beta)}{\partial \theta_{jk} \partial \phi_{lm}} \right) \quad k=1,\dots,q, m=1,\dots,p \\ (F_{\Theta_j\Theta_l})_{km} &= F(\theta_{jk}, \theta_{lm}) = -E \left(\frac{\partial^2 \log l(\beta)}{\partial \theta_{jk} \partial \theta_{lm}} \right) \quad k=1,\dots,q, m=1,\dots,q, j,l=1,\dots,d. \end{aligned} \quad (5.5.6)$$

pour calculer les éléments de F introduisons le processus $\{Z_t, t \in \mathbb{Z}\}$ autorégressif d -périodique défini par $\Phi_i(L)\theta_i(L)Z_{i+d\tau} = \varepsilon_{i+d\tau}$, $i=1,\dots,d$, $\tau \in \mathbb{Z}$

ou encore

$$\Phi_i(L)Z_{i+d\tau} = \varepsilon_{i+d\tau}, \quad i=1,\dots,d, \tau \in \mathbb{Z} \quad (5.5.7)$$

où $\Phi_i(L) = 1 - \sum_{j=1}^p \phi_{ij}L^j$, $\phi_{ij} = \sum_{k=0}^j -\phi_{ik}\theta_{i,j-k}$ et $P = p+q$ avec la convention que

$\phi_{ik}, \theta_{ik} = 0$ si $k > q$ ou $k < 0$. Nous allons exprimer les éléments de F en terme des autocovariances $\gamma_h^{(i)}$ du processus $\{Z_t, t \in \mathbb{Z}\}$, d'où

Proposition 5.1

La matrice d'information de Fisher pour le paramètre vectoriel, β , d'un modèle autorégressif moyenne mobile périodique est symétrique, diagonale par blocs et ces éléments sont donnés par les équations suivantes :

$$F(\phi_{jk}, \phi_{lm}) = \delta_{jl} \sigma_j^{-2} \sum_{s=0}^q \sum_{h=0}^q \theta_{js} \theta_{lh} \gamma_{(m+h)-(k+s)}^{(j-(k+s))}, \quad j,l=1,\dots,d, k,m=1,\dots,p \quad (5.5.8)$$

$$F(\theta_{jk}, \theta_{lm}) = \delta_{jl} \sigma_j^{-2} \sum_{s=0}^p \sum_{h=0}^p \phi_{js} \phi_{lh} \gamma_{(m+h)-(k+s)}^{(j-(k+s))}, \quad j,l=1,\dots,d, k,m=1,\dots,q \quad (5.5.9)$$

$$F(\theta_{jk}, \phi_{lm}) = -\delta_{jl} \sigma_j^{-2} \sum_{s=0}^p \sum_{h=0}^q \phi_{js} \theta_{lh} \gamma_{(m+h)-(k+s)}^{(j-(k+s))}, \quad j,l=1,\dots,d, k=1,\dots,q, m=1,\dots,p, \quad (5.5.10)$$

$$\text{où } \delta_{jl} = \begin{cases} 1 & \text{si } j = l \\ 0 & \text{sinon} \end{cases}.$$

Comme pour le cas du modèle ARMA classique, voir Zahaf (1998), pour améliorer le comportement et la convergence de l'algorithme nous introduisons dans le schémas précédent la prévision de l'erreur notée $\bar{\varepsilon}_t$ et une suite (γ) de scalaires positifs tendant vers 0 telle que $(\sum \gamma_t)$ diverge. La structure des relation de récurrences est modifiée pour mieux faire apparaître le caractère récurrent de l'algorithme

Proposition 5.4.2

Sous les mêmes hypothèses A et B, les équations de (5.4.18) peuvent être remplacées par les équations suivantes

$$\begin{cases} \varepsilon_{i+d\tau} = y_{i+d\tau} - \hat{\beta}_{d\tau}^T \bar{\varphi}_{i+d\tau}, \\ \hat{\beta}_{i+d\tau} = \hat{\beta}_{i+d\tau-1} + \gamma_{i+d\tau} R_{i+d\tau}^{-1} \psi_{i+d\tau} \varepsilon_{i+d\tau}, \\ \bar{\varepsilon}_{i+d\tau} = y_{i+d\tau} - \hat{\beta}_{i+d\tau-1}^T \bar{\varphi}_{i+d\tau}, \\ \psi_{i+1+d\tau} = \hat{\theta}_{i1}^{t-1} \psi_{i+d\tau} + \hat{\theta}_{i2}^{t-1} \psi_{i-1+d\tau} + \dots + \hat{\theta}_{iq}^{t-1} + \bar{\varphi}_{i+d\tau-q+1}, \\ R_{i+1+d\tau} = R_{i+d\tau} + \gamma_{i+d\tau} [\psi_{i+1+d\tau} \psi_{i+1+d\tau}^T - R_{i+d\tau}] \end{cases} \quad (5.4.19)$$

pour, $\tau \in \mathbb{Z}$. $i = 1, \dots, d$ où $R_{i+d\tau} = \gamma_{i+d\tau} \bar{R}_{i+d\tau}$ et $\bar{\varphi}_t$ est obtenu à partir de φ_t selon (5.4.0) en remplaçant ε_t par $\bar{\varepsilon}_t$.

Preuve

A partir de (5.2.4) et (5.2.6) nous pouvons écrire

$$y_{i+d\tau} - \phi_{i1} y_{i+d\tau-1} - \phi_{i2} y_{i+d\tau-2} - \dots - \phi_{ip} y_{i+d\tau-p} = \hat{\varepsilon}_{i+d\tau}(\beta) - \theta_{i1} \hat{\varepsilon}_{i+d\tau-1}(\beta) - \dots - \theta_{iq} \hat{\varepsilon}_{i+d\tau-q}(\beta) \quad i=1,2,\dots,d, \quad (5.5.11)$$

où $\hat{\varepsilon}_t(\beta)$ est l'erreur de prévision

Cette dernière équation peut s'écrire pour le vecteur d'observations $y^T = (y_1, y_2, \dots, y_{Nd})$ comme suit,

$$Ay + B\hat{\varepsilon} = CI_* \quad (5.5.12)$$

où $\hat{\varepsilon} = (\hat{\varepsilon}_1, \hat{\varepsilon}_2, \dots, \hat{\varepsilon}_{Nd})^T$ et $I_* = (y_{1-p}, y_{2-p}, \dots, y_0, \hat{\varepsilon}_{1-q}, \hat{\varepsilon}_{2-q}, \dots, \hat{\varepsilon}_0)^T$ et A, B, C sont des matrices dont l'expression est donnée par (voir Vecchia (1985a) pour le cas de modèles moyenne mobile périodique).

$$A_{ij} = \begin{cases} -[\phi_{i,i-j}] & \text{si } j < i \\ 1 & \text{si } j = i \\ 0 & \text{si } j > i \end{cases} \quad i, j = 1, \dots, Nd$$

$$B_{ij} = \begin{cases} -[\theta_{i,i-j}] & \text{si } j < i \\ 1 & \text{si } j = i \\ 0 & \text{si } j > i \end{cases} \quad i, j = 1, \dots, Nd$$

$$C_{ij} = \begin{cases} [\phi_{i,p+i-j}] & \text{si } j \leq i \text{ et } j \leq p \\ -[\theta_{i,p+q+i-j}] & \text{si } i \leq j-p \text{ et } j > p \\ 0 & \text{ailleurs} \end{cases} \quad i = 1, \dots, Nd, \quad j = 1, \dots, p+q$$

où

$$[\phi_{ij}] = \begin{cases} \phi_{ij} & \text{si } j = 1, \dots, p \\ 0 & \text{ailleurs} \end{cases}, \quad [\theta_{ij}] = \begin{cases} \theta_{ij} & \text{si } j = 1, \dots, q \\ 0 & \text{ailleurs} \end{cases}$$

Pour N assez grand, nous pouvons considérer la fonction de vraisemblance conditionnelle à $I_* = 0$. Les $\{\hat{\varepsilon}_{i+d\tau}(\beta)\}$ seront donc Gaussiennes et non corrélées. La densité conjointe $f(\cdot)$ du vecteur ε sera donc donnée par

$$f(\varepsilon) = (1/\sqrt{2\pi})^{Nd/2} \left(\prod_{t=1}^{Nd} \sigma_t^2 \right)^{-1/2} \exp - \sum_{t=1}^{Nd} \frac{\hat{\varepsilon}_t^2(\beta)}{2\sigma_t^2}. \quad (5.5.13)$$

la matrice Jacobienne de la transformation de R^{Nd} qui applique y sur $\hat{\varepsilon}$ à l'aide de (5.5.12) est triangulaire inférieure et les éléments diagonaux sont égaux à 1 (à cause de la forme particulière des matrices A, B, C). Par conséquent la fonction densité de y conditionnellement à $I_* = 0$ sera donnée par (5.5.13). La fonction de vraisemblance conditionnelle $l(\beta)$ est par conséquent : $l(\beta) = f(y) = f(\hat{\varepsilon})$ où $f(y)$ est la densité conjointe de y . $\hat{\varepsilon}$ s'exprime par rapport à β et y au moyen de l'équation (5.5.12). Ainsi,

$$\begin{aligned} \log l(\beta) &= -Nd/2 \log 2\pi - 1/2 \sum_{t=1}^{Nd} \log \sigma_t^2 - 1/2 \sum_{t=1}^{Nd} \frac{\hat{\varepsilon}_t^2(\beta)}{\sigma_t^2} \\ &= -Nd/2 \log 2\pi - N/2 \sum_{i=1}^d \log \sigma_i^2 - 1/2 \sum_{\tau=0}^{N-1} \sum_{i=1}^d \frac{\hat{\varepsilon}_{i+d\tau}^2(\beta)}{\sigma_i^2} \end{aligned}$$

d'où

$$\begin{aligned} \frac{\partial^2 \log l(\beta)}{\partial \beta \partial \beta^T} &= -\frac{1}{2} \frac{\partial^2}{\partial \beta \partial \beta^T} \left[\sum_{t=1}^{Nd} \frac{\hat{\varepsilon}_t^2(\beta)}{\sigma_t^2} \right] \\ &= -\frac{1}{2} \sum_{t=1}^{Nd} \frac{2}{\sigma_t^2} \left[\frac{\partial \hat{\varepsilon}_t(\beta)}{\partial \beta} \frac{\partial \hat{\varepsilon}_t(\beta)}{\partial \beta^T} + \hat{\varepsilon}_t(\beta) \frac{\partial^2 \hat{\varepsilon}_t(\beta)}{\partial \beta \partial \beta^T} \right]. \end{aligned}$$

$$\text{or, } E \left(\hat{\varepsilon}_t(\beta) \frac{\partial^2 \hat{\varepsilon}_t(\beta)}{\partial \beta \partial \beta^T} \right) = 0.$$

En effet, la dérivée seconde de $\hat{\varepsilon}_t(\beta)$ par rapport à β s'exprime comme combinaison linéaire de seulement du passé de y_t , comme $\hat{\varepsilon}_t(\beta)$ est non corrélée avec le passé de y_t et donc elle le sera aussi avec la dérivée seconde de $\hat{\varepsilon}_t(\beta)$. Ainsi, la matrice F s'exprime aussi par

$$\begin{aligned} NF &= \sum_{t=1}^{Nd} \frac{1}{\sigma_t^2} E \left(\frac{\partial \hat{\varepsilon}_t(\beta)}{\partial \beta} \frac{\partial \hat{\varepsilon}_t(\beta)}{\partial \beta^T} \right), \\ &= \sum_{\tau=0}^{N-1} \sum_{i=1}^d \frac{1}{\sigma_i^2} E \left(\frac{\partial \hat{\varepsilon}_{i+d\tau}(\beta)}{\partial \beta} \frac{\partial \hat{\varepsilon}_{i+d\tau}(\beta)}{\partial \beta^T} \right). \end{aligned}$$

En exploitant la d -périodicité du modèle i.-e.

$$E\left(\frac{\partial \hat{\varepsilon}_{i+d\tau}(\beta)}{\partial \beta} \frac{\partial \hat{\varepsilon}_{i+d\tau}(\beta)}{\partial \beta^T}\right) = E\left(\frac{\partial \hat{\varepsilon}_i(\beta)}{\partial \beta} \frac{\partial \hat{\varepsilon}_i(\beta)}{\partial \beta^T}\right) \text{ pour } \tau = 0, \dots, N-1, i = 1, \dots, d,$$

nous obtenons

$$F = \sum_{i=1}^d \frac{1}{\sigma_i^2} E\left(\frac{\partial \hat{\varepsilon}_i(\beta)}{\partial \beta} \frac{\partial \hat{\varepsilon}_i(\beta)}{\partial \beta^T}\right). \quad (5.5.14)$$

Calculons maintenant, pour $i = 1, \dots, d$, $\tau \in Z$, la quantité :

$$E\left(\frac{\partial \hat{\varepsilon}_{i+d\tau}(\beta)}{\partial \beta} \frac{\partial \hat{\varepsilon}_{i+d\tau}(\beta)}{\partial \beta^T}\right)$$

L'équation (5.5.11) peut se mettre également sous la forme condensée suivante

$$\phi_i(L)y_{i+d\tau} = \theta_i(L)\hat{\varepsilon}_{i+d\tau}(\beta) \quad i = 1, \dots, d$$

où L est l'opérateur retard et les polynômes :

$$\phi_i(L) = 1 - \phi_{i1}L - \phi_{i2}L^2 - \dots - \phi_{ip}L^p \quad \text{et} \quad \theta_i(L) = 1 - \theta_{i1}L - \theta_{i2}L^2 - \dots - \theta_{iq}L^q \quad \text{sont supposés}$$

minimales et de racines distinctes.

Les dérivées de $\hat{\varepsilon}_i(\beta)$ par rapport à β s'expriment différemment selon que nous dérivons par rapport à ϕ_{jk} ou à θ_{lm} , $j, l = 1, \dots, d$ $k = 1, \dots, p$ $m = 1, \dots, q$, ainsi

$$\frac{\partial \hat{\varepsilon}_{i+d\tau}(\beta)}{\partial \phi_{jk}} = \begin{cases} 0 & \text{Si } j \neq i \\ -L^k \theta_i(L)y_{i+d\tau} = -L^k \phi_i^{-1}(L)\hat{\varepsilon}_{i+d\tau}(\beta) & \text{Si } j = i \end{cases} \quad (5.5.15)$$

et

$$\frac{\partial \hat{\varepsilon}_{i+d\tau}(\beta)}{\partial \theta_{lm}} = \begin{cases} 0 & \text{Si } j \neq i \\ L^m \theta_i^{-2}(L)\phi_i(L)y_{i+d\tau} = L^m \theta_i^{-1}(L)\hat{\varepsilon}_{i+d\tau}(\beta) & \text{Si } l = i \end{cases} \quad (5.5.16)$$

Par conséquent les éléments de la matrice F se calculent à partir de (5.5.14). Plus précisément,

$$F(\phi_{jk}, \phi_{lm}) = \sum_{i=1}^d \frac{1}{\sigma_i^2} E\left(\frac{\partial \hat{\varepsilon}_i(\beta)}{\partial \phi_{jk}} \frac{\partial \hat{\varepsilon}_i(\beta)}{\partial \phi_{lm}}\right)$$

$$F(\phi_{jk}, \phi_{lm}) = \begin{cases} 0 & \text{Si } j \neq l \\ \frac{1}{\sigma_j^2} E\left(\frac{\partial \hat{\varepsilon}_j(\beta)}{\partial \phi_{jk}} \frac{\partial \hat{\varepsilon}_j(\beta)}{\partial \phi_{jm}}\right) & \text{Si } j = l \end{cases} \quad \begin{matrix} j, l = 1, \dots, d, \\ k, m = 1, \dots, p. \end{matrix}$$

En utilisant (5.5.7), (5.5.15) et (5.5.16) il vient que

$$F(\phi_{jk}, \phi_{lm}) = \begin{cases} 0 & \text{Si } j \neq l \\ \frac{1}{\sigma_j^2} E(L^k \theta_j(L)Z_j L^m \theta_j(L)Z_j) & \text{Si } j = l \end{cases}$$

$$= \delta_{jl} \sigma_j^{-2} E\left(\left(\sum_{s=0}^q \theta_{js} L^{s+k}\right) Z_j \left(\sum_{h=0}^q \theta_{jh} L^{h+m}\right) Z_j\right),$$

$$= \delta_{jl} \sigma_j^{-2} \sum_{s=0}^q \sum_{h=0}^q \theta_{js} \theta_{jh} E(Z_{j-(k+s)} Z_{j-(m+h)}),$$

$$= \delta_{jl} \sigma_j^{-2} \sum_{s=0}^q \sum_{h=0}^q \theta_{js} \theta_{jh} \gamma_{(m+h)-(k+s)}^{(j-(k+s))},$$

où $\gamma_h^{(i)} = \gamma(i, i-h)$ est la fonction d'autocovariance de la période i et à l'horizon h du processus $\{Z_t, t \in \mathbb{Z}\}$ défini par (5.5.7).

De même

$$F(\theta_{jk}, \phi_{lm}) = \sum_{i=1}^d \frac{1}{\sigma_i^2} E \left(\frac{\partial \hat{\varepsilon}_i(\beta)}{\partial \theta_{jk}} \frac{\partial \hat{\varepsilon}_i(\beta)}{\partial \phi_{lm}} \right)$$

$$F(\theta_{jk}, \phi_{lm}) = \begin{cases} 0 & \text{Si } j \neq l \\ \frac{1}{\sigma_j^2} E \left(\frac{\partial \hat{\varepsilon}_j(\beta)}{\partial \theta_{jk}} \frac{\partial \hat{\varepsilon}_j(\beta)}{\partial \phi_{jm}} \right) & \text{Si } j = l \end{cases} \quad \begin{matrix} j, l = 1, \dots, d, \\ k = 1, \dots, q, m = 1, \dots, p. \end{matrix}$$

En se servant toujours de (5.5.7), (5.5.15) et (5.5.16)

$$F(\theta_{jk}, \theta_{lm}) = \begin{cases} 0 & \text{Si } j \neq l \\ \frac{1}{\sigma_j^2} E(L^k \phi_j(L) Z_j L^m \phi_j(L) Z_j) & \text{Si } j = l \end{cases}$$

$$= \delta_{jl} \sigma_j^{-2} E \left(\left(\sum_{s=0}^p \phi_{js} L^{s+k} \right) Z_j \left(\sum_{h=0}^p \phi_{jh} L^{h+m} \right) Z_j \right),$$

$$= \delta_{jl} \sigma_j^{-2} \sum_{s=0}^p \sum_{h=0}^p \phi_{js} \phi_{jh} E(Z_{j-(k+s)} Z_{j-(m+h)}),$$

$$= \delta_{jl} \sigma_j^{-2} \sum_{s=0}^p \sum_{h=0}^p \phi_{js} \phi_{jh} \gamma_{(m+h)-(k+s)}^{(j-(k+s))},$$

Finalement,

$$F(\theta_{jk}, \theta_{lm}) = \sum_{i=1}^d \frac{1}{\sigma_i^2} E \left(\frac{\partial \hat{\varepsilon}_{i+d\tau}(\beta)}{\partial \theta_{jk}} \frac{\partial \hat{\varepsilon}_{i+d\tau}(\beta)}{\partial \theta_{lm}} \right)$$

$$F(\theta_{jk}, \theta_{lm}) = \begin{cases} 0 & \text{Si } j \neq l \\ \frac{1}{\sigma_j^2} E \left(\frac{\partial \hat{\varepsilon}_j(\beta)}{\partial \theta_{jk}} \frac{\partial \hat{\varepsilon}_j(\beta)}{\partial \theta_{jm}} \right) & \text{Si } j = l \end{cases} \quad \begin{matrix} j, l = 1, \dots, d, \\ k, m = 1, \dots, q. \end{matrix}$$

d'où

$$F(\theta_{jk}, \phi_{lm}) = \begin{cases} 0 & \text{Si } j \neq l \\ \frac{1}{\sigma_j^2} E(-L^k \phi_j(L) Z_j L^m \theta_j(L) Z_j) & \text{Si } j = l \end{cases}$$

$$= -\delta_{jl} \sigma_j^{-2} E \left(\left(\sum_{s=0}^p \phi_{js} L^{s+k} \right) Z_j \left(\sum_{h=0}^q \theta_{jh} L^{h+m} \right) Z_j \right),$$

$$= -\delta_{jl} \sigma_j^{-2} \sum_{s=0}^p \sum_{h=0}^q \phi_{js} \theta_{jh} E(Z_{j-(k+s)} Z_{j-(m+h)}),$$

$$= -\delta_{jl} \sigma_j^{-2} \sum_{s=0}^p \sum_{h=0}^q \phi_{js} \theta_{jh} \gamma_{(m+h)-(k+s)}^{(j-(k+s))},$$

■

Il semble que les équations (5.5.8), (5.5.9) et (5.5.10) permettent d'obtenir les éléments de la matrice, cependant il reste à évaluer les autaucovariances $\gamma_h^{(i)}$ du processus autorégressif périodique $\{Z_t, t \in \mathbb{Z}\}$ sur la base des paramètres. Nous présenterons ce calcul dans la section 5.4.

5.2 Algorithme RML_{MZ}

Comme pour le cas d'un modèle ARMA classique, et pour essayer d'améliorer l'algorithme RML périodique, nous pensons à prolonger l'algorithme RML_{MZ} (qui est une amélioration de RML dans le cas d'un modèle ARMA) aux modèles PARMA périodiques. Nous remplaçons donc la matrice des dérivées secondes \bar{R}_t par son espérance mathématique.

Nous savons que, dans le cas gaussien, la fonction critère $V_t(\beta) = \frac{1}{2} \sum_{k=1}^t (1/\sigma_k^2) \hat{\varepsilon}_k^2(\beta)$ est identique (à une constante près) à l'opposé du logarithme de la fonction de vraisemblance conditionnelle. Donc la matrice d'information de Fisher, $I_t(\beta)$ relative à une série y^t est égale à l'espérance de la dérivée seconde, par rapport à β , de la fonction critère $V_t(\beta)$. Ainsi, en gardant le même principe, nous remplacerons dans les équations de récurrences (5.4.18) la matrice des dérivées secondes \bar{R}_t par la matrice d'information de Fisher associée aux paramètres du modèle PARMA (relativement à la série de taille t).

La justification de cette amélioration est la suivante : D'après Mélard (1989), la matrice \bar{R}_t , qui représente une approximation de la matrice de covariances asymptotique, est fort variable et cela introduit des perturbations nuisibles à l'estimation courante. Zahaf (1998) a remplacé cette quantité par son espérance mathématique qui n'est rien d'autre que la matrice d'information de Fisher. Par ailleurs nous savons que pour un modèle ARMA classique la matrice des variances-covariance de l'estimateur du maximum de vraisemblance tend, à une constante multiplicative près (Taille de l'observation), vers l'inverse de la matrice d'information, ce résultat relatif aux modèles ARMA classique n'est pas toujours valable pour les modèles ARMA_t à coefficients évolutifs dans le temps voir Mélard (1985b). Cependant Pagano (1978) a montré que la matrice de covariances de l'estimateur des moments (qui est équivalent à l'estimateur du maximum de vraisemblance dans le cas de ces modèles PAR_d) d'un modèle autorégressif périodique pur converge presque sûrement, à une constante multiplicative près, qui est la taille d'un canal N , vers l'inverse de la matrice d'information de Fisher d'un modèle PAR_d. Il a donné ensuite l'expression de cette matrice. Pour ces raisons

nous avons pensé à remplacer, comme pour le cas classique, la matrice des dérivées secondes \bar{R}_t par la matrice d'information de Fisher d'un modèle PARMA périodique.

Comme $F_N(\beta) = N.F$ (voir Whittle, (1953)) dans le cas ARMA classique et $F_{Nd}(\beta) = N.F$ (voir Pagano, (1978)) donc nous remplaçons dans les équations (5.4.28) \bar{R}_{t-1}^{-1} par $((t-1)/d)F^{-1}(\hat{\beta}_{t-1})$.

Ainsi nous aurons

$$\hat{\beta}_{i+d\tau} = \hat{\beta}_{i+d\tau-1} + \hat{\sigma}_i^{-2(\tau)} ((i+d\tau-1)/d) F_{i+d\tau}^{-1} \psi_{i+d\tau} \hat{\varepsilon}_{i+d\tau}; \quad (5.5.17) \quad (5.2.1)$$

Le calcul récursif de \bar{R}_t est cependant inutile, il sera donc supprimé, d'où

Proposition 5.2

Les équations qui suit, fournit un estimateur pour les paramètre dont la fonction critère est optimum

$$\left\{ \begin{array}{l} \varepsilon_{i+d\tau} = y_{i+d\tau} - \hat{\beta}_{d\tau}^T \bar{\varphi}_{i+d\tau}, \\ F_{i+d\tau} = F(\hat{\beta}_{i+d\tau-1}) \\ \hat{\beta}_{i+d\tau} = \hat{\beta}_{i+d\tau-1} + \sigma_i^{2(\tau)} ((i+d\tau-1)/d) F_{i+d\tau}^{-1} \psi_{i+d\tau} \varepsilon_{i+d\tau} \\ \bar{\varepsilon}_{i+d\tau} = y_{i+d\tau} - \hat{\beta}_{i+d\tau-1}^T \bar{\varphi}_{i+d\tau}, \\ \psi_{i+1+d\tau} = \hat{\theta}_{i1}^{t-1} \psi_{i+d\tau} + \hat{\theta}_{i2}^{t-1} \psi_{i-1+d\tau} + \dots + \hat{\theta}_{iq}^{t-1} + \bar{\varphi}_{i+d\tau-q+1}, \end{array} \right. \quad (5.5.18)(5.2.2)$$

pour $i = 1, \dots, d$, $\tau \in \mathbb{Z}$.

où $\sigma_i^{2(\tau)}$ sera obtenue à partir de (5.3.11).

5.3 Propriétés asymptotiques des estimateurs RML et RML_{MZ} adaptables au cas de modèles périodiques

Notons que les algorithmes RML et RML_{MZ} périodiques ont la même forme que ceux du cas classique. C'est à dire qu'ils peuvent être exprimés en termes d'équation différentielle déterministes. Pour étudier en détails le comportement asymptotique des estimateurs obtenus par ces algorithmes nous pouvons renvoyer le lecteur à Ljung et Soderström (1983) pour le RML et à Zahaf (1998) pour le RML_{MZ}.

5.4. Calcul des autocovariances d'un modèle autorégressif périodique

Dans le paragraphe qui suit sera présentée une procédure pour un calcul des autocovariances d'un modèle AR périodique pur. Notons qu'il est possible d'estimer celles-ci au moyen des équations de Yule-Walker périodiques, étant donné que les paramètres du modèle sont connus (estimés au fur et à mesure).

Le processus $\{Z_{i+d\tau}, i = 1, \dots, d; \tau \in \mathbb{Z}\}$ est représenté par le modèle AR_d(P), le nous avons alors :

$$Z_{i+d\tau} - \sum_{j=1}^P \phi_{ij} Z_{i+d\tau-j} = \varepsilon_{i+d\tau}, \quad i = 1, \dots, d; \tau \in \mathbb{Z}. \quad (5.5.19)$$

Notons par ϕ_i , $i = 1, \dots, d$ le $P \times 1$ vecteur des paramètres pour la saison i définit par $\phi_i = (\phi_{i1}, \phi_{i2}, \dots, \phi_{iP})^T$. Multiplions (5.5.19) par $Z_{i+d\tau-h}$, $h = 1, \dots, P$, et prenons l'espérance mathématique des deux membres alors nous obtenons pour chaque i l'ensemble des $P+1$ équations qui s'écrit de la manière suivante :

$$\left\{ \begin{array}{l} \Gamma_i \phi_i = \gamma_i \\ \sigma_i^2 = \gamma_0^{(i)} - \phi_i^T \gamma_i \end{array} \right. \quad i = 1, \dots, d \quad (5.5.20)$$

avec : $(\Gamma_i)_{kj} = \gamma_{j-k}^{(i)}$, $(k, j = 1, \dots, P)$ et $\gamma_i = (\gamma_1^{(i)}, \gamma_2^{(i)}, \dots, \gamma_P^{(i)})^T$. (Pour plus de détails voir Vecchia (1985b)).

L'équation (5.5.20) permet d'obtenir les paramètres du modèle à partir des autocovariances et vice-versa. La détermination des paramètres par le biais de (5.5.20) nécessite la résolution de d systèmes indépendants de taille P . Cela requiert un nombre d'opérations (multiplications et divisions) de l'ordre de $O(d(P)^3)$. Pour le problème inverse c'est à dire la détermination des $\gamma_h^{(i)}$ sur la base de ϕ , Nous généralisons, aux modèles AR périodiques, la procédure de McLeod, (1975) relative aux modèles ARMA. Le système (5.5.20) sera donc transformé au système suivant :

$$G\gamma = g \quad (5.5.21)$$

où $\gamma = (\gamma_0^{(1)}, \gamma_1^{(1)}, \dots, \gamma_P^{(1)}; \gamma_0^{(2)}, \dots, \gamma_P^{(2)}; \dots; \gamma_1^{(d)}, \dots, \gamma_P^{(d)})^T$ est le $(P+1)d$ -vecteur des autocovariances, G est une matrice carrée d'ordre $(P+1)d$ qui est fonction de Φ et dont l'expression sera donnée en fonction du modèle choisit. g est un vecteur de dimension $(P+1)d$, définit pour $i = 1, \dots, (P+1)d$

$$g(i) = \begin{cases} \sigma_{k+1}^2 & \text{si } i = kd + 1 \\ 0 & \text{sinon} \end{cases} \quad k = 0, \dots, d-1$$

La résolution du système (5.5.21) fournit les $\gamma_h^{(i)}$ $i = 1, \dots, d$, $h = 0, \dots, P$, (dont nous ne aurons besoin que pour $h = 0, \dots, P-1$) avec un nombre d'opérations de l'ordre de $O(((P+1) \times d)^3)$.

Remarques

Notons que, par définition, la matrice F est une fonction du vecteur des paramètres β , or d'après les relation (5.5.8)-(5.5.10) celle-ci est exprimée en fonction des autocovariances $\gamma_h^{(i)}$ d'un certain processus. C'est pour cela que nous avons donné ci-dessus un algorithme de calcul de $\gamma_h^{(i)}$ en fonction de β . Or la complexité de l'expression de la matrice G et le nombre d'opération nous incite à utiliser l'approximation suivante :

Pour estimer les autocovariances $\gamma_h^{(i)}$ nous utiliserons les autocovariances empiriques exprimées sous forme récursive.

Notons par $\tilde{\gamma}_h^{(i,\tau)}$ la l'autocovariance empirique à l'horizon h , obtenue par une observation de taille $t(i, \tau) = i + d\tau$ alors nous avons

$$\begin{aligned} \tilde{\gamma}_h^{(i,\tau)} &= \frac{1}{\tau+1} \sum_{k=0}^{\tau} y_{i+dk} y_{i+dk-h} \\ &= \frac{1}{\tau+1} \left[\sum_{k=0}^{\tau-1} y_{i+dk} y_{i+dk-h} + y_{i+d\tau} y_{i+d\tau-h} \right] \end{aligned}$$

d'où

$$\tilde{\gamma}_h^{(i,\tau)} = \frac{\tau}{\tau+1} \tilde{\gamma}_h^{(i,\tau-1)} + \frac{1}{\tau+1} y_{i+d\tau} y_{i+d\tau-h} \quad (5.5.22)$$

5.5. Exemples

Considérons le modèle $AR_2(2)$ suivant :

$$y_{i+2\tau} - \phi_{i1} y_{i+2\tau-1} - \phi_{i2} y_{i+2\tau-2} = \varepsilon_{i+2\tau} \quad i = 1, 2 \quad \tau \in \mathbb{Z}.$$

où $\beta^T = (\phi_{11}, \phi_{12}, \phi_{21}, \phi_{22})$ est le vecteur des paramètres. Alors la matrice d'information de Fisher est diagonale par bloc, et elle s'écrit comme suit :

$$F = \begin{pmatrix} F_{11} & 0 \\ 0 & F_{22} \end{pmatrix} \quad \text{où } (F_{jj})_{km} = F(\phi_{jk}, \phi_{jm}) \quad j = 1, 2 \quad k, m = 1, 2. \quad \text{D'après (5.5.8), nous avons}$$

donc,

$$F(\phi_{jk}, \phi_{jm}) = \frac{1}{\sigma_j^2} (\theta_{j0} \theta_{j0} \gamma_{j-k}^{(m-k)}) \quad j = 1, 2 \quad k, m = 1, 2. \quad \text{où } \gamma_h^{(i)} = \gamma(i, i-h) \text{ est la fonction}$$

d'autocovariance de la saison i et à l'horizon h du processus $\{Z_t, t \in \mathbb{Z}\}$ définit par :

$$\phi_i(L)\theta_i(L)Z_{i+d\tau} = \varepsilon_{i+d\tau}, \quad i=1,2, \quad \tau \in \mathbb{Z}$$

avec $\theta_i(L) \equiv 1$, ou encore

$$\Phi_i(L)Z_{i+d\tau} = \varepsilon_{i+d\tau}, \quad i=1,2, \quad \tau \in \mathbb{Z}$$

où $\Phi_i(L) = \phi_i(L) = 1 - \sum_{j=1}^2 \phi_{ij}L^j$, $\phi_{ij} = \phi_{ij}$ $i=1,2$ $j=1,2$ et $P=p=2$, $q=0$ avec la convention que $\phi_{ik}, \theta_{ik} = 0$ si $k > q$ ou $k < 0$ et $\theta_{i0} = \phi_{i0} = -1$. L'expression de la matrice est par conséquent,

$$F = \begin{pmatrix} \frac{1}{\sigma_1^2} \begin{pmatrix} \gamma_0^{(2)} & \gamma_1^{(2)} \\ \gamma_1^{(2)} & \gamma_0^{(1)} \end{pmatrix} & 0 \\ 0 & \frac{1}{\sigma_2^2} \begin{pmatrix} \gamma_0^{(1)} & \gamma_1^{(1)} \\ \gamma_1^{(1)} & \gamma_0^{(2)} \end{pmatrix} \end{pmatrix}$$

Calculons maintenant les autocovariances $\gamma_h^{(i)}$, $i=1,2$ $h=0,1$

Les équations de Yule-Walker périodiques relatif au processus $\{Z_t, t \in \mathbb{Z}\}$ donnent

$$\begin{cases} \gamma_1^{(1)}\phi_{11} + \gamma_2^{(1)}\phi_{12} = \gamma_0^{(1)} - \sigma_1^2 \\ \gamma_0^{(2)}\phi_{11} + \gamma_1^{(2)}\phi_{12} = \gamma_1^{(1)} \\ \gamma_1^{(2)}\phi_{11} + \gamma_0^{(1)}\phi_{12} = \gamma_2^{(1)} \\ \gamma_1^{(2)}\phi_{21} + \gamma_2^{(2)}\phi_{22} = \gamma_0^{(2)} - \sigma_2^2 \\ \gamma_1^{(1)}\phi_{21} + \gamma_1^{(1)}\phi_{12} = \gamma_1^{(2)} \\ \gamma_1^{(1)}\phi_{21} + \gamma_0^{(2)}\phi_{22} = \gamma_2^{(2)} \end{cases}$$

Ce système sera transformé au système suivant

$$\begin{bmatrix} 1 & -\phi_{11} & -\phi_{12} & 0 & 0 & 0 \\ 0 & 1 & 0 & -\phi_{11} & -\phi_{12} & 0 \\ -\phi_{12} & 0 & 1 & 0 & -\phi_{11} & 0 \\ 0 & 0 & 0 & 1 & -\phi_{21} & -\phi_{22} \\ -\phi_{21} & -\phi_{22} & 0 & 0 & 1 & 0 \\ 0 & -\phi_{21} & 0 & -\phi_{22} & 0 & 1 \end{bmatrix} \begin{bmatrix} \gamma_0^{(1)} \\ \gamma_1^{(1)} \\ \gamma_2^{(1)} \\ \gamma_0^{(2)} \\ \gamma_1^{(2)} \\ \gamma_2^{(2)} \end{bmatrix} = \begin{bmatrix} \sigma_1^2 \\ 0 \\ 0 \\ \sigma_2^2 \\ 0 \\ 0 \end{bmatrix}$$

La résolution de ce dernier fournira les autocovariances recherchées.

2- Etant donné le modèle MA₃(2) suivant :

$$y_{i+3\tau} = \varepsilon_{i+3\tau} - \theta_{i1}\varepsilon_{i+3\tau-1} - \theta_{i2}\varepsilon_{i+3\tau-2} \quad i=1,2,3 \quad \tau \in \mathbb{Z}.$$

où $\beta^T = (\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}, \theta_{31}, \theta_{32})$ est le vecteur des paramètres. Alors la matrice d'information de Fisher est diagonale par bloc, et elle s'écrit comme suit :

$$F = \begin{pmatrix} F_{11} & 0 & 0 \\ 0 & F_{11} & 0 \\ 0 & 0 & F_{11} \end{pmatrix}. \quad \text{où } (F_{jj})_{km} = F(\theta_{jk}, \theta_{jm}) \quad j=1,2,3 \quad k,m=1,2. \quad \text{D'après (5.5.9),}$$

nous avons,

$$F(\phi_{jk}, \phi_{jm}) = \frac{1}{\sigma_j^2} (\phi_{j0}\phi_{j0}\gamma_{j-k}^{(m-k)}) \quad j=1,2,3 \quad k,m=1,2. \quad \text{où } \gamma_h^{(i)} = \gamma(i, i-h) \text{ est la fonction}$$

d'autocovariance de la saison i et à l'horizon h du processus $\{Z_t, t \in \mathbb{Z}\}$ définit par :

$$\phi_i(L)\theta_i(L)Z_{i+d\tau} = \varepsilon_{i+d\tau}, \quad i=1,2,3, \quad \tau \in \mathbb{Z} \text{ avec } \phi_i(L) \equiv 1$$

ou encore

$$\Phi_i(L)Z_{i+d\tau} = \varepsilon_{i+d\tau}, \quad i = 1,2,3, \tau \in \mathbb{Z}$$

où $\Phi_i(L) = 1 - \sum_{j=1}^2 \phi_{ij} L^j$, $\phi_{ij} = \theta_{ij}$ $i = 1,2,3$ $j = 1,2$ et $P = q = 2$ avec la convention que

$$\phi_{ik}, \theta_{ik} = 0 \quad \text{si } k > q \quad \text{ou } k < 0 \quad \text{et } \theta_{i0} = \phi_{i0} = -1.$$

Dans ce cas, la matrice d'information de Fisher relative aux paramètres β est donnée par :

$$F = \begin{pmatrix} \frac{1}{\sigma_1^2} \begin{pmatrix} \gamma_0^{(3)} & \gamma_1^{(3)} \\ \gamma_1^{(3)} & \gamma_0^{(2)} \end{pmatrix} & 0 & 0 \\ 0 & \frac{1}{\sigma_2^2} \begin{pmatrix} \gamma_0^{(1)} & \gamma_1^{(1)} \\ \gamma_1^{(1)} & \gamma_0^{(3)} \end{pmatrix} & 0 \\ 0 & 0 & \frac{1}{\sigma_3^2} \begin{pmatrix} \gamma_0^{(2)} & \gamma_1^{(2)} \\ \gamma_1^{(2)} & \gamma_0^{(1)} \end{pmatrix} \end{pmatrix}.$$

3- Soit le modèle PARMA₂(1,1) suivant :

$$y_{i+2\tau} - \phi_{i1} y_{i+2\tau-1} = \varepsilon_{i+2\tau} - \theta_{i1} \varepsilon_{i+2\tau-1} \quad i = 1,2 \quad \tau \in \mathbb{Z}.$$

où $\beta^T = (\phi_{11}, \theta_{11}, \phi_{21}, \theta_{21})$ est le vecteur des paramètres. Alors la matrice d'information de Fisher est diagonale par bloc, et elle s'écrit comme suit :

$$F = \begin{pmatrix} F_{11} & 0 \\ 0 & F_{22} \end{pmatrix}. \quad \text{D'après (5.5.8), nous avons donc,}$$

$$F(\phi_{jk}, \phi_{jm}) = \frac{1}{\sigma_j^2} \left(\sum_{s=0}^1 \sum_{h=0}^1 \theta_{js} \theta_{jh} \gamma_{h-s}^{(-s)} \right) \quad j = 1,2 \quad k, m = 1. \quad \text{où } \gamma_h^{(i)} = \gamma(i, i-h) \text{ est la fonction}$$

d'autocovariance d'ordre i et à l'horizon h du processus $\{Z_t, t \in \mathbb{Z}\}$ définit par :

$$\phi_i(L)\theta_i(L)Z_{i+d\tau} = \varepsilon_{i+d\tau}, \quad i = 1,2, \tau \in \mathbb{Z}$$

ou encore

$$\Phi_i(L)Z_{i+d\tau} = \varepsilon_{i+d\tau}, \quad i = 1,2, \tau \in \mathbb{Z}$$

où $\Phi_i(L) = 1 - \sum_{j=1}^2 \phi_{ij} L^j$, on trouve $\phi_{i1} = \sum_{k=0}^1 -\phi_{ik} \theta_{i,1-k} = \theta_{i1} + \phi_{i1}$, $i = 1,2$ et

$$\phi_{i2} = \sum_{k=0}^2 (-\phi_{ik} \theta_{i,2-k}) = -\phi_{i1} \theta_{i1}, \quad \text{avec la convention que } \phi_{ik}, \theta_{ik} = 0 \quad \text{si } k > q \quad \text{ou } k < 0 \quad \text{et}$$

$\theta_{i0} = \phi_{i0} = -1$. L'expression de la matrice est par conséquent,

$$F(\phi_{11}, \phi_{11}) = \sigma_1^{-2} [\gamma_0^{(2)} - 2\theta_{11} \gamma_1^{(2)} + (\theta_{11})^2 \gamma_0^{(1)}]$$

$$F(\theta_{11}, \theta_{11}) = \sigma_1^{-2} [\gamma_0^{(2)} - 2\phi_{11} \gamma_1^{(2)} + (\phi_{11})^2 \gamma_0^{(1)}]$$

$$F(\theta_{11}, \phi_{11}) = F(\phi_{11}, \theta_{11}) = \sigma_1^{-2} [\gamma_0^{(2)} - (\phi_{11} + \theta_{11}) \gamma_1^{(2)} + (\phi_{11} \theta_{11}) \gamma_0^{(1)}]$$

$$F(\phi_{21}, \phi_{21}) = \sigma_2^{-2} [\gamma_0^{(2)} - 2\theta_{21} \gamma_1^{(2)} + (\theta_{21})^2 \gamma_0^{(1)}]$$

$$F(\theta_{21}, \theta_{21}) = \sigma_2^{-2} [\gamma_0^{(2)} - 2\phi_{21} \gamma_1^{(2)} + (\phi_{21})^2 \gamma_0^{(1)}]$$

$$F(\theta_{21}, \phi_{21}) = F(\phi_{21}, \theta_{21}) = \sigma_2^{-2} [\gamma_0^{(2)} - (\phi_{21} + \theta_{21}) \gamma_1^{(2)} + (\phi_{21} \theta_{21}) \gamma_0^{(1)}].$$

Quant aux autocovariances elles seront obtenues par résolution du système de l'exemple 1.

Chapitre 6

Implémentation

1. Introduction

Aux chapitres 4 et 5, nous avons présenté les algorithmes RLS, RML et RML_{MZ} appliqués aux modèles ARMA et aux modèles PARMA. Nous avons donné la formulation générale et discuté la convergence mais sans pour autant abordé l'aspect pratique de l'implémentation, en effet, la simplicité des algorithmes récurrents étudiés ne doit pas faire oublier qu'en pratique, leur implémentation numérique n'est pas une tâche aisée. Et ce, est dû au fait que les procédures d'optimisations requises sont locales, ce qui rend l'algorithme sensible aux valeurs initiales, choix de la suite de gains scalaires, choix de la suite de gains matriciels et à la stabilité de la valeur courante. Il s'agit alors de prendre quelques précautions pour réduire cet effet et éviter que l'estimateur $\hat{\beta}_t$ ne s'éloigne de la vraie valeur du paramètre β . Nous discuterons donc le choix des valeurs initiales, le problème de la stabilité numérique et l'admissibilité de la valeur courante de l'estimateur (dans notre cas le modèle PARMA doit être causal et inversible, cf. chapitre 2), de l'utilisation d'un facteur d'oubli et enfin des problèmes posés par la matrice de gain R_t .

2. Choix des conditions expérimentales

2.1. Choix des valeurs initiales

Les algorithmes récurrents, comme ils tous sont basés sur des routines d'optimisation numériques, sont très sensibles aux valeurs initiales. Il faut donc choisir de bonnes valeurs de départ afin d'assurer la convergence dans des temps raisonnables. Pour empêcher que l'estimateur ne soit très sensible aux premières valeurs de la série, il est important de bien choisir la valeur initiale R_0 . Notons que par exemple pour l'algorithme RLS la matrice R_t , qui s'interprète comme étant la matrice de corrélation empirique du vecteur φ_t , ne pourra être inversible (compte tenu des valeurs initiales nulles) que si $t \geq p$. Il faudrait donc calculer au départ R_p et $\hat{\beta}_p$ associés, comme valeurs initiales de la récurrence de l'algorithme. En pratique, des valeurs convenables pour R_0 et $\hat{\beta}_0$ seront

$$R_0 = \frac{1}{M} I_p \quad \text{et} \quad \hat{\beta}_0 = 0 \quad \text{avec } M \text{ est un nombre positif assez grand.}$$

Par exemple en prenant l'algorithme RLS, ceci revient à remplacer la suite de matrices R_t par la suite

$$R_t = \frac{1}{t} \left[\sum_{k=1}^t \varphi_k \varphi_k^T + \frac{1}{M} I_p \right].$$

Nous remarquons que lorsque t s'accroît indéfiniment l'influence de la perturbation initiale sera négligeable. De plus cette méthode garantit la propriété d'inversibilité de toutes les matrices R_t (car c'est une suite croissante de matrices vérifiant : $R_{t+1} \geq R_t \geq \frac{1}{M} I_p > 0$).

Si nous n'avons aucune information sur β , alors nous choisirons $\hat{\beta}_0 = 0$, et pour $R_0 = \mathcal{G}^{-1} I_d$ où \mathcal{G} est un nombre assez grand.

Si nous avons un estimateur de β (notée $\hat{\beta}^0$), nous poserons $\hat{\beta}_0 = \hat{\beta}^0$ et si nous avons une grande confiance en la valeur initiale de $\hat{\beta}_0$, alors nous choisirons

$$R_0 = \text{cov} [\hat{\beta}_0] / E[y_t^2].$$

2.2. Stabilité et admissibilité de la valeur courante $\hat{\beta}_t$

Pour chaque valeur t , nous devons contrôler l'admissibilité des valeurs $\hat{\beta}_t$. C'est à dire que les polynômes autorégressifs et moyenne mobile doivent vérifier les conditions de stationnarité et d'inversibilité données dans le premier et deuxième chapitre. Ljung et Söderström (1983) ont proposé différentes approches qui consistent à tester d'abord l'admissibilité et en cas de rejet à appliquer une projection de l'estimateur obtenu dans un ensemble D_M où les valeurs sont acceptables. Nous avons appliqué la méthode suivante pour projeter $\hat{\beta}_t$ dans l'ensemble D_M :

1. calcul de $\hat{\beta}_t = \hat{\beta}_{t-1} + \mathcal{G}_t$,
2. Nous testons si $\hat{\beta}_t$ non admissible,
3. dans ce cas posons $\mathcal{G}_t^1 = \rho \mathcal{G}_t$,
4. Nous testons une autre fois si $\hat{\beta}_t = \hat{\beta}_{t-1} + \mathcal{G}_t^1$
5. si $\hat{\beta}_t$ est encore non admissible alors Nous posons : $\mathcal{G}_t^2 = \rho \mathcal{G}_t^1$,
6. Nous continuons de la même manière jusqu'à obtenir une valeur admissible pour $\hat{\beta}_t$.

Cette méthode nécessite le choix de ρ , généralement $\rho = 0.5$ convient mais nous avons observé que pour des modèles p et/ou q sont grand il faut parfois diminuer cette valeur. Notons que pour la classe de modèles PARMA le test d'admissibilité (causalité et inversibilité) de l'estimation courante peut être obtenu soit en utilisant la technique de Gladyshev (1961) ou bien par le biais de la formule de Bentarzi et Hallin (1994).

2.3. Choix de suite de gains scalaires et facteur d'oubli λ_t

L'algorithme récurrent est très sensible aux premières valeurs de l'estimateur. Pour remédier à cet inconvénient Ljung et Söderström (1983) propose d'incorporer un facteur d'oubli, une suite $(\lambda_t)_{t \in \mathbb{N}}$ dans l'algorithme qui permet d'atténuer les variations des paramètres. La suite $(\lambda_t)_{t \in \mathbb{N}}$ sera reliée à la suite $(\gamma)_{t \in \mathbb{N}}$ par la relation suivante :

$$\lambda_t = \frac{\gamma_{t-1}}{\gamma_t} [1 - \gamma_t]$$

où (γ) est une suite de scalaires positifs tendant vers 0 telle que $(\sum \gamma_t)$ diverge, en pratique, on utilise souvent

$$\lambda_t = \lambda^0 \lambda_{t-1} + (1 - \lambda^0)$$

Avec les conditions initiales suivantes :

$$\lambda_0 = 0.95 \text{ et } \lambda^0 = 0.99$$

2.4. Choix de la suite de gains matriciels R_t

Nous devons imposer que R_t soit symétrique définie positive Ljung et Söderström (1983) proposent un algorithme utilisant la *factorisation* U-D (Bierman (1977)) de p_t , définie par $p_t = \gamma_t R_t^{-1}$.

Du fait que p_t est symétrique définie positive on peut écrire que $p_t = U_t D_t U_t^T$, où U_t est une matrice triangulaire supérieure avec tous les éléments diagonaux égaux à 1 et D_t est une matrice diagonale. Cette décomposition pour le calcul de p_t garantit le caractère symétrique définie positive de p_t .

3. Simulation et validation

Pour vérifier l'efficacité de la méthode d'estimation proposée, nous proposons d'appliquer cette méthode à des données artificielles considérées comme réalisation d'un processus PARMA $_d(p,q)$, (ARMA pour $d = 1$) complètement spécifié (i-e : les paramètres du modèle sont parfaitement connus). Pour cela nous allons utiliser la simulation. Les simulations sont souvent utilisées pour l'analyse, lorsqu'il est difficile d'obtenir des résultats théoriques. Ici elles nous serviront à confirmer les résultats obtenus.

3.1. Choix des modèles

Nous avons tenu compte des points suivants :

- Choix des modèles simulés :

Nous avons choisi une classe assez large de modèles. Nous avons pris des valeurs de p , q et d assez grandes. (nous avons fait varier p et q entre 0 et 4 et d entre 1 et 12)

- Choix des paramètres :

Les paramètres sont choisis de sorte à vérifier l'hypothèse sur le modèle c'est à dire que les polynômes autorégressifs et moyennes mobiles ont des racines distinctes et de plus vérifient les conditions de stationnarité et d'inversibilité.

- Taille des séries :

Nous avons fait varier N entre 100 et 1000 afin d'observer le comportement des estimateurs et tenir compte des réalités économiques.

- Nombre de répétitions :

REP = 1000, afin d'obtenir une meilleure précision des estimateurs obtenus.

- Les sorties de simulation contiennent les moyennes des valeurs estimées des paramètres $\hat{\beta}_t$, les erreurs-types et la somme des carrés.

3.2. Description du programme de simulation

Pour chaque modèle spécifié, nous avons conçu une procédure propre à celui-ci, nous allons décrire le moule général de ces programmes. Le programme comporte principalement :

- 1- Une première boucle sur le nombre de répétitions $i = 1, \text{REP}$,
- 2- Une seconde boucle sur le temps $t = 1, N$. A l'intérieur de cette boucle, nous pouvons considérer globalement trois parties :

- (a) La première partie consiste à simuler les séries y_t :
Nous générons un processus bruit blanc gaussien ε_t .

Nous déterminons alors la valeur courante de la série y_t en fonction des ε_t et des valeurs des paramètres du modèle à simuler.

- (b) la seconde partie consiste à estimer le vecteur des paramètres β à chaque instant t .
Pour éviter les effets néfastes des valeurs initiales, nous débuterons l'estimation à partir de la 51^{ème} observation ($t = 51$) et nous n'employons pas les 50 premières données.
- (c) la troisième partie consiste à remettre à jour les différentes statistiques utiles pour voir le comportement de la méthode.

3.3. Détail des simulations

Les programmes ont été écrit en Matlab 5.0. Les simulations ont été effectuées sur un Personal Computer avec un processeur Pentium I cadencé à 200 Mhz. Nous avons développé les différentes procédures intervenant dans nos programmes de simulation.

3.3.1. Génération de séries artificielles

Pour générer une série artificielle de taille N , nous procédons selon les étapes suivantes :

- 1- Nous générons un nombre aléatoire appartenant à la loi uniforme au moyen d'une commande prédéterminée dans l'environnement Matlab 5.0. En utilisant la formule de Box-Muller nous générons ainsi une réalisation d'une variable normale dont les paramètres sont spécifiés dans le programme. Nous pouvons ensuite générer le processus bruit blanc.
- 2- Calculer la valeur courante de la série notée y_t pour p et q fixés, ainsi que pour des valeurs des paramètres $\phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q$ fixés à priori (modèle que l'on souhaite simuler).

3.3.2. Estimation récurrente des paramètres

L'initialisation des différentes valeurs intervenant dans l'estimation récurrente se trouve au début de la boucle sur les répétitions. Nous commençons à partir de la 51^{ème} observation.

Nous aurons besoin au sein de cette procédure d'évaluer l'inverse d'une matrice (la matrice des dérivées secondes R) à chaque instant t cette tâche est accomplie par l'intermédiaire d'une procédure de l'environnement Matlab.

3.3.3. Calcul des statistiques de comparaison

Nous obtenons, à chaque répétition, un vecteur de valeurs de l'estimateur dont les composantes représentent ces valeurs pour chaque temps, Nous calculons ainsi les moyennes et les écart-types pour toutes les répétitions pour voir le comportement de la méthode. Pour les séries simulées, les moyennes et les écart-types des valeurs obtenues des estimateurs sont donnés sous forme graphique, où l'abscisse est le nombre d'observations. Cela permet

d'observer le comportement dans le temps de l'estimateur récurrent et des différentes statistiques.

4. Résultats

Nous avons simulé, pour différents modèles $\text{PARMA}_d(p,q)$ où p et q varient entre 0 et 4, et d entre 1 et 12, 1000 séries de longueur comprise entre 100 et 1000.

Nous nous sommes d'abord intéressés à la précision des estimateurs obtenus. Pour cela nous avons présenté sur les graphiques les moyennes (respectivement les écart-types) en fonction du nombre d'observations. Ces graphiques sont notés $*_{\gamma.pqd.j}$, selon la convention suivante : * est **M** pour les moyennes (sur les 1000 séries simulées) des estimateurs et **ET** pour les écart-types, le symbole ? est **M** pour distinguer l'algorithme RML, **Z** pour l'algorithme RML_{MZ} et **C** pour un graphique de comparaison entre les deux algorithmes, pqd représente l'ordre et la période du modèle estimé, j (facultatif) est par contre le numéro de l'occurrence du graphique d'un même modèle.

Nous allons maintenant, après avoir présenté quelques-uns des modèles simulés et estimés, donner quelques commentaires sur les graphiques.

4.1. Résultats concernant les modèles ARMA classiques

- 1- **Modèle ARMA(1,0)** d'équation, $(1-0.5L)y_t = \varepsilon_t$ la racines du polynôme est $a_1 = 2$, la valeur initiale de l'estimateur qui a été prise est $\hat{\phi}_0 = 0$. Le graphique **M_M101** indique que l'estimateur récurrent converge rapidement vers la solution et qu'à partir de la 40^{ème} observation l'erreur commise est de l'ordre de 0.05. Cependant sur le graphique **ET_M101** on aperçoit que l'écart-type décroît vers 0 mais en présentant un biais de 0.1.
- 2- **Modèle ARMA(0,1)** d'équation, $y_t = (1-0.5L)\varepsilon_t$ la racines du polynôme est $b_1 = 2$, la valeur initiale de l'estimateur qui a été prise est $\hat{\theta}_0 = 0$. Le graphique **M_M011** indique que l'estimateur récurrent converge rapidement vers la solution et qu'à partir de la 60^{ème} observation l'erreur commise est de l'ordre de 0.05. Notons que l'estimateur RML est plus précis pour les modèles autorégressifs contrairement aux modèles moyennes mobiles et ce est dû à l'approximation utilisées. Pour l'écart-type, nous pouvons observer le même comportement.
- 3- **Modèle ARMA(1,1)** Nous avons choisi l'équation $(1+0.5L)y_t = (1-0.5L)\varepsilon_t$ les racines des polynômes sont $b_1=-2$, pour l'autorégressif et $b_2 = 2$, pour le polynôme moyenne mobile. Nous observons sur le graphique **M_M111** que les estimations des deux paramètres se rapprochent de plus en plus des vraies valeurs de ceux-ci et ce à partir de la 40^{ème} observation, nous remarquons une légère supériorité dans la précision pour le cas de l'estimation du paramètre autorégressif.
- 4- **Modèle ARMA(2,0)** L'équation du modèle choisit est $(1-0.4L+0.7L^2)y_t = \varepsilon_t$ notons que ce modèle tel qu'il a été choisit est causal. Nous remarquons sur le graphique **M_M201** que les estimations autorégressives sont stables, autours des vraies valeurs des paramètres, à partir de la 35^{ème} observation. Les erreurs-types, quant à elles, sont au voisinage de 0.1 à partir de la 40^{ème} observation mais en présentant un biais de 0.07.

- 5- **Modèle ARMA(0,2)** Nous avons choisit le modèle suivant $y_t = (1 - 0.5L - 0.3L^2)\varepsilon_t$, les paramètres sont choisis de sorte que la condition d'inversibilité soit vérifiée nous avons commencé à partir des valeurs initiales nulles les résultats obtenus par les graphiques \mathbf{M}_M201 et \mathbf{ET}_M201 montre que l'algorithme donne de bonnes estimations qu'à partir de la 60^{ème} valeurs, cela est expliqué par le fait que les modèles à composantes moyenne mobile présentent une imprécision, par rapport aux modèles autorégressifs, due à la non-linéarité de la dérivée de la fonction critère.
- 6- **Modèle ARMA(2,1)** L'instance sélectionnée du modèle ARMA(2,1) est celui choisit par Zahaf (1998) d'équation $(1 - 0.3L + 0.02L^2)y_t = (1 + 0.05L)\varepsilon_t$. Notons que ce modèle est causal est inversible, nous avons commencé l'algorithme RML par les valeurs initiales suivantes $\hat{\phi}_1^{(0)} = 0.15$ $\hat{\phi}_2^{(0)} = -0.005$ $\hat{\theta}_1^{(0)} = -0.025$. Nous remarquons sur le graphique \mathbf{ET}_M211 que les écart-types présentent des biais de 0.06, par contre dans \mathbf{M}_M211 les estimations obtenues convergent, vers les vraies valeurs des paramètres, à partir de la 55^{ème} observation et que l'erreur commise est de l'ordre de 0.05.
- 7- **Modèle ARMA(3,0)** Nous avons opté, cette fois-ci, pour le modèle causal d'équation suivante $(1 + 0.5L - 0.3L^2 - 0.5L^3)y_t = \varepsilon_t$. Les valeurs initiales ont été choisies nulles, dans le graphe \mathbf{M}_M301 on aperçoit la précision de l'estimateur de l'ordre de 0.05 et ce à partir de la 50^{ème} observation, ce qui est confirmé dans le graphique \mathbf{ET}_M301 , en effet les erreurs-types convergent en présentant un biais de 0.06.
- 8- **Modèle ARMA(0,3)** Le modèle particulier choisit est d'équation stochastique suivante $y_t = (1 - 0.5L + 0.3L^2 + 0.5L^3)\varepsilon_t$ avec $\sigma_i^2 = 1$, $i = 1, 2, 3$. Nous avons constater à partir du graphique \mathbf{M}_M013 que les estimations sont proches des valeurs réelles des paramètres à partir de la 60^{ème} valeur avec une précision de 0.05 (voir aussi graphique \mathbf{ET}_M301).
- 9- **Modèle ARMA(4,0)** Nous avons choisit le modèle d'équation stochastique donnée comme suit : $(1 - 0.5L + 0.5L^2 + 0.3L^3)y_t = \varepsilon_t$. Les valeurs initiales sont choisies nulles, les paramètres sont dans le domaine des valeurs admissibles. Nous avons remarqué sur les graphiques $\mathbf{M}_M401.1$, $\mathbf{M}_M401.2$, $\mathbf{ET}_M401.1$ et $\mathbf{ET}_M401.2$ une stabilité de l'estimation avec une erreur de 0.05 vers la vraie valeur du paramètre et ce à partir de la 50^{ème} observation.
- 10- **Modèle ARMA(2,2)** Dans le présent cas nous avons choisit le modèle, causal et inversible, d'équation stochastique suivante : $(1 - 0.4L + 0.55L^2)y_t = (1 + 0.4L - 0.5L^2)\varepsilon_t$, avec $\sigma_i^2 = 1$, $i = 1, 2$. Nous avons observé sur les graphiques $\mathbf{M}_M221.1$, $\mathbf{M}_M221.2$, $\mathbf{ET}_M221.1$ et $\mathbf{ET}_M221.2$ que les estimations produites par l'algorithme RML convergent vers les vraies valeurs des paramètres à partir de la 60^{ème} observation avec une précision de 0.06.

4.2. Résultats concernant les modèles PARMA périodiques

4.2.1 Modèle $\mathbf{PAR}_2(1)$

Nous avons choisit le modèle d'équation
$$\begin{cases} (1 - 0.5L)y_{1+2\tau} = \varepsilon_{1+2\tau} \\ (1 + 0.5L)y_{2+2\tau} = \varepsilon_{2+2\tau} \end{cases}$$
 où $\{\varepsilon_t\}$ est un bruit blanc

périodique de variance unité. Notons que les paramètres choisis vérifient bien les conditions de causalité et d'inversibilité. En commençant par les valeurs initiales nulles nous remarquons que l'algorithme RML commence à se stabiliser à partir de la 60^{ème} observation et que l'erreur commise est de l'ordre de 0.01.

Le graphique ET_M102 indique que les erreurs types tendent vers zéro mais en présentant un biais de 0.1.

En utilisant les mêmes paramètres nous remarquons sur le graphique M_Z102 que l'algorithme RML_{MZ} donne des estimations plus précises, et que l'algorithme RML présente un biais.

4.2.2 Modèle PAR₃(1)

Nous avons choisit le modèle d'équation
$$\begin{cases} (1 - 0.7L)y_{1+3\tau} = \varepsilon_{1+3\tau} \\ (1 - 0.3L)y_{2+3\tau} = \varepsilon_{2+3\tau} \\ (1 + 0.4L)y_{3+3\tau} = \varepsilon_{3+3\tau} \end{cases}$$
 de variance unité.

Il est claire que les paramètres choisis sont dans le domaine des valeurs admissibles. Les valeurs initiales sont choisies nulles. L'algorithme RML, comme nous observons dans le graphique M_M103, commence à approcher la vraie valeur à partir de la 50^{ème} observation avec une précision de 0.04. Le graphique ET_M103 indique aussi que les écarts type des estimations sont proches de zéro.

4.2.3 Modèle PAR₄(1)

Le modèle utilisé dans le présent cas est celui d'équation suivante
$$\begin{cases} (1 - 1.5L)y_{1+3\tau} = \varepsilon_{1+3\tau} \\ (1 + 0.5L)y_{1+3\tau} = \varepsilon_{1+3\tau} \\ (1 - 1.2L)y_{1+3\tau} = \varepsilon_{1+3\tau} \\ (1 - 0.7L)y_{1+3\tau} = \varepsilon_{1+3\tau} \end{cases}$$

Les variances sont choisies égales à un. Nous avons commencé par des valeurs nulles. Il est à noter que les paramètres choisis sont dans le domaine des valeurs admissibles. L'algorithme RML, comme nous observons dans le graphique M_M104, tend à approcher la vraie valeur de chaque paramètre à partir de la 50^{ème} observation avec une précision de 0.05. Le graphique ET_M104 indique aussi que les écarts type des estimations sont proches de zéro.

4.2.4 Modèle PAR₇(1)

Nous avons choisit le modèle d'équation suivante
$$\begin{cases} (1 - 1.5L)y_{1+3\tau} = \varepsilon_{1+3\tau} \\ (1 + 0.5L)y_{1+3\tau} = \varepsilon_{1+3\tau} \\ (1 - 1.2L)y_{1+3\tau} = \varepsilon_{1+3\tau} \\ (1 - 0.7L)y_{1+3\tau} = \varepsilon_{1+3\tau} \\ (1 - 0.8L)y_{1+3\tau} = \varepsilon_{1+3\tau} \\ (1 + 1.4L)y_{1+3\tau} = \varepsilon_{1+3\tau} \\ (1 - 0.9L)y_{1+3\tau} = \varepsilon_{1+3\tau} \end{cases}$$

Les variances sont choisies égales à un. Nous avons commencé par des valeurs nulles. Il est à noter que les paramètres choisis sont dans le domaine des valeurs admissibles. L'algorithme RML, comme nous observons dans le graphique M_M107 , tend à approcher la vraie valeur de chaque paramètre à partir de la 60^{ème} observation avec une précision de 0.05. Le graphique ET_M107 indique aussi que les écarts type des estimations sont proches de zéro.

4.2.5 Modèle $PAR_2(2)$

Cette fois-ci nous avons choisit à appliquer l'algorithme RML sur le modèle suivant

$$\begin{cases} (1 - 0.6L - 0.3L^2)y_{1+2\tau} = \varepsilon_{1+2\tau} \\ (1 + 0.2L + 0.4L^2)y_{2+2\tau} = \varepsilon_{2+2\tau} \end{cases} .$$

Nous avons également choisit les paramètres de sorte à ce que le modèle soit causal, les valeurs initiales sont choisies nulles, à partir de l'ordre de 65 observations l'estimateur prend une tendance stable avec une erreur de 0.01 comme l'indiquent les graphiques M_M202 et ET_M202 .

4.2.6 Modèle $PMA_2(1)$

Nous avons choisit le modèle (de Cipra, (1985)) d'équation $\begin{cases} (1 - 0.6L)\varepsilon_{1+2\tau} = y_{1+2\tau} \\ (1 - 0.3L)\varepsilon_{2+2\tau} = y_{2+2\tau} \end{cases}$ il est

clair que les paramètres choisis vérifient les conditions de causalité et d'inversibilité, les valeurs initiales sont choisies nulles. Nous remarquons que l'algorithme RML commence à se stabiliser à partir de la 70^{ème} observation et que l'erreur commise est de l'ordre de 0.07. Alors que l'algorithme RML_{MZ} est nettement plus stable que son homologue, en effet la précision dans ce cas est de l'ordre de 0.01.

4.2.7 Modèle $PMA_3(1)$

Le modèle utilisé dans le présent cas est celui d'équation suivante $\begin{cases} (1 - 0.6L)\varepsilon_{1+2\tau} = y_{1+2\tau} \\ (1 - 0.3L)\varepsilon_{1+2\tau} = y_{1+2\tau} \\ (1 + 0.5L)\varepsilon_{1+2\tau} = y_{1+2\tau} \end{cases}$

Les variances sont choisies égales à un. Le modèle ainsi choisit est inversible, notons que le graphique M_M013 et ET_M013 montrent qu'à partir de la 70^{ème} valeur les estimations sont proches de la vraie valeur avec une erreur de 0.01.

4.2.8 Modèle $PMA_4(1)$

L'algorithme RML est appliqué sur le modèle suivant $\begin{cases} (1 - 0.6L)\varepsilon_{1+2\tau} = y_{1+2\tau} \\ (1 - 0.5L)\varepsilon_{1+2\tau} = y_{1+2\tau} \\ (1 - 0.3L)\varepsilon_{1+2\tau} = y_{1+2\tau} \\ (1 + 0.5L)\varepsilon_{1+2\tau} = y_{1+2\tau} \end{cases}$ avec des

variances du bruit unité. Le modèle ainsi utilisé est inversible. Nous pouvons observer les

mêmes résultats que pour le modèle $PAR_4(1)$ à part une légère imprécision relativement aux modèles moyennes mobiles.

4.2.9 Modèle $PMA_2(2)$

L'algorithme RML est appliqué sur le modèle suivant $\begin{cases} (1 - 0.6L + 0.3L^2)\varepsilon_{1+2\tau} = y_{1+2\tau} \\ (1 + 0.5L - 0.3L^2)\varepsilon_{2+2\tau} = y_{2+2\tau} \end{cases}$ où les

innovations sont supposées de variance un. Nous avons commencé l'algorithme avec des valeurs initiales nulles, nous avons observé qu'à partir la 60^{ème} observation l'estimateur tend à se stabiliser autour de la vraie valeur. Notons aussi que d'après le graphique ET_{M022} , l'estimateur RML semble présenter un léger biais.

4.2.10 Modèle $PARMA_2(1,1)$

Nous avons choisit le modèle suivant (Vecchia, (1985b))

$$\begin{cases} (1 + 0.8L)y_{1+2\tau} = (1 - 0.9L)\varepsilon_{1+2\tau} \\ (1 - 0.6L)y_{2+2\tau} = (1 + 0.5L)\varepsilon_{2+2\tau} \end{cases} .$$

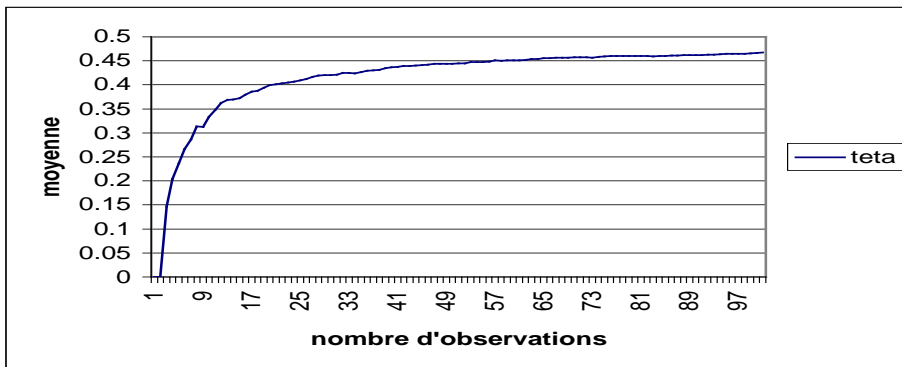
Notons que pour ce genre de modèles, plus complexe que

les modèles PAR purs et PMA purs, la précision est plus difficile à obtenir et de bonnes valeurs initiales sont nécessaires. Nous avons choisit donc les valeurs initiales suivantes :

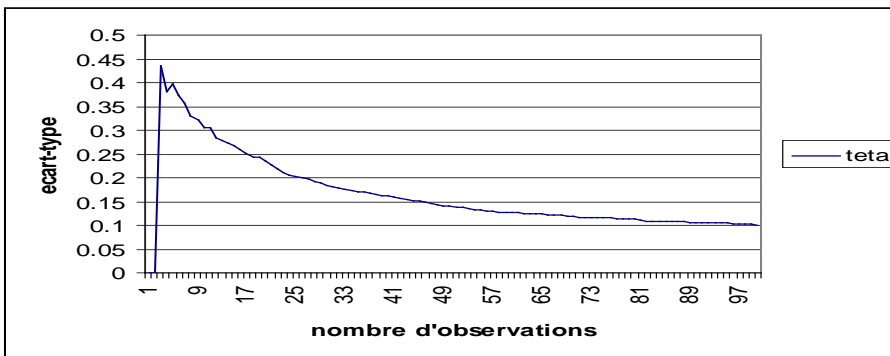
$\hat{\phi}_{11}^{(0)} = 0.2$, $\hat{\theta}_{11}^{(0)} = -0.4$, $\hat{\phi}_{21}^{(0)} = 0.3$ et $\hat{\theta}_{21}^{(0)} = -0.5$. Nous observons sur le graphique M_M112 que les estimateurs se stabilisent à partir de la 80^{ème} valeurs (pour quelques paramètres la précision est meilleur).

Résultats concernant le cas de modèles ARMA classiques

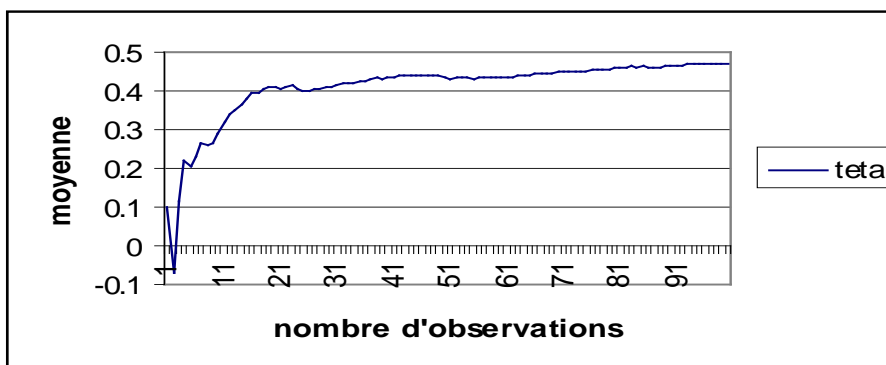
Modèle ARMA(0,1)



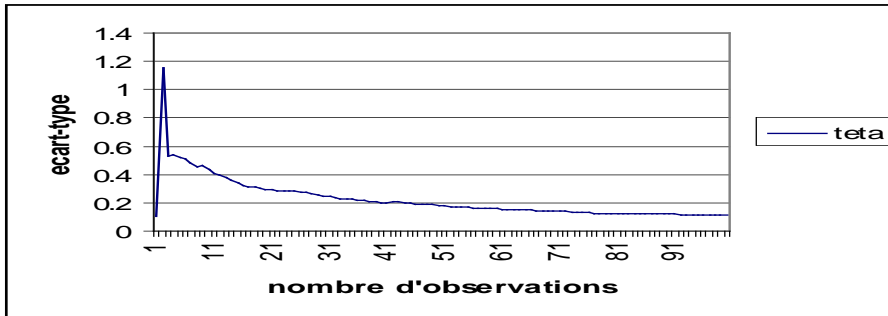
Graphique M_M011



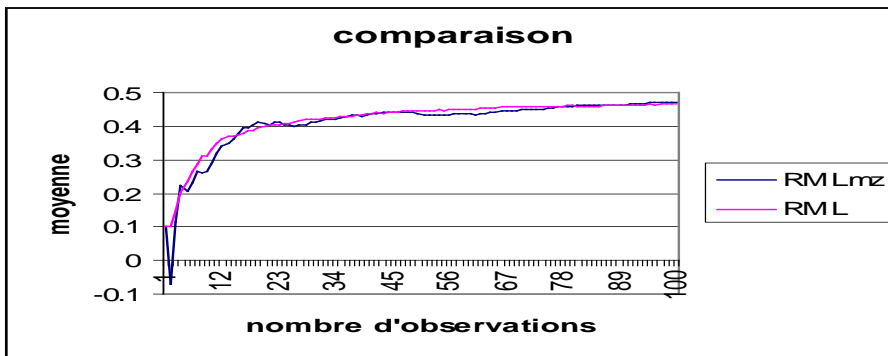
Graphique ET_M011



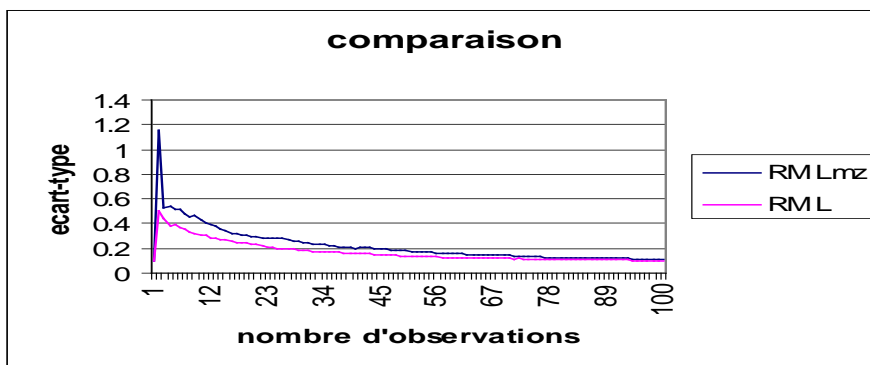
Graphique M_Z011



Graphique ET_Z011

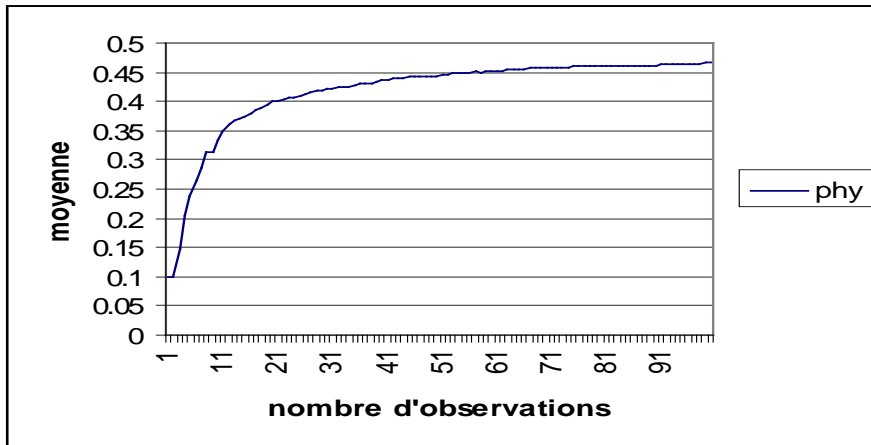


Graphique C_M011

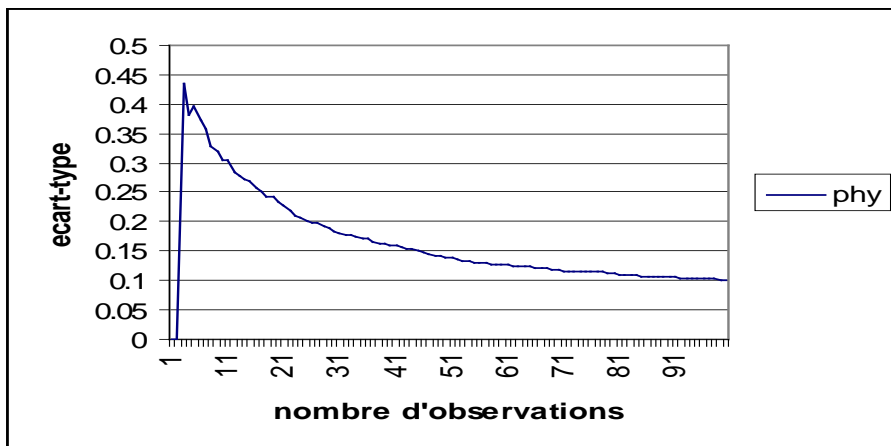


Graphique C_M011

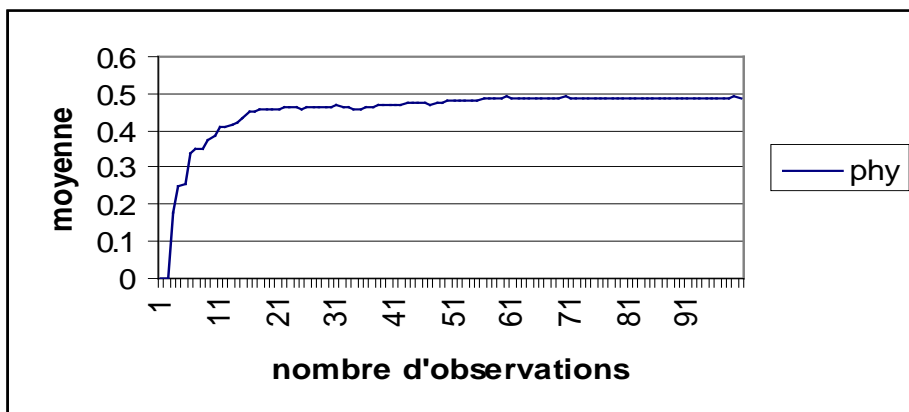
Modèle ARMA(1,0)



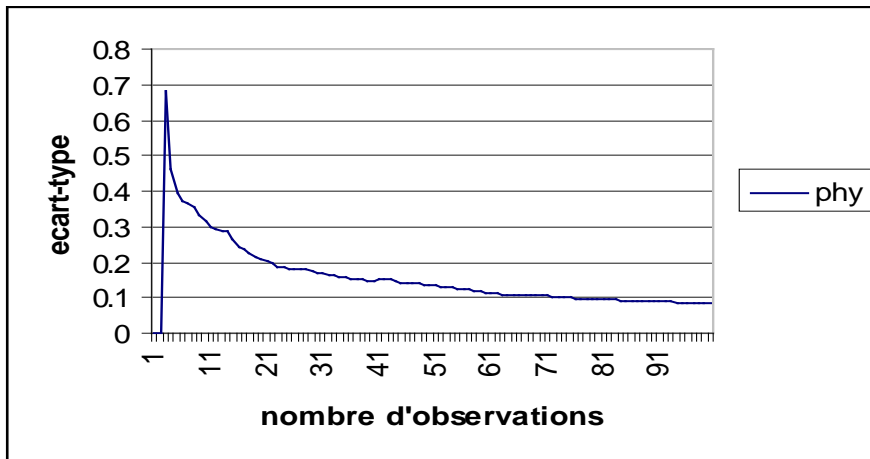
Graphique M_M101



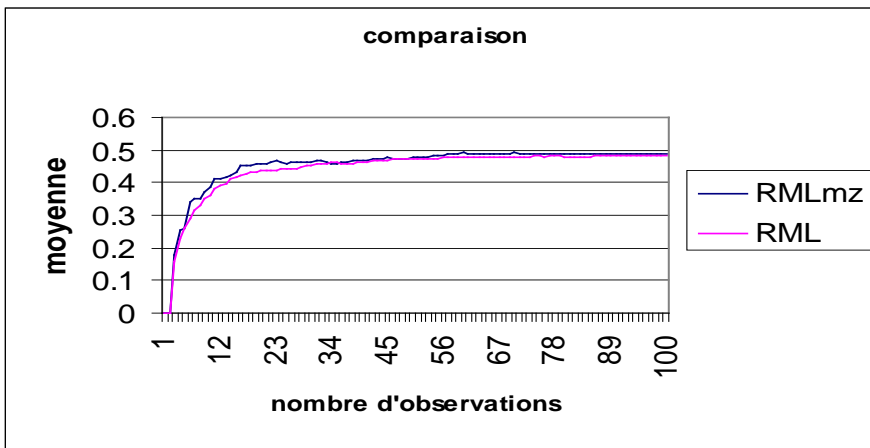
Graphique ET_M011



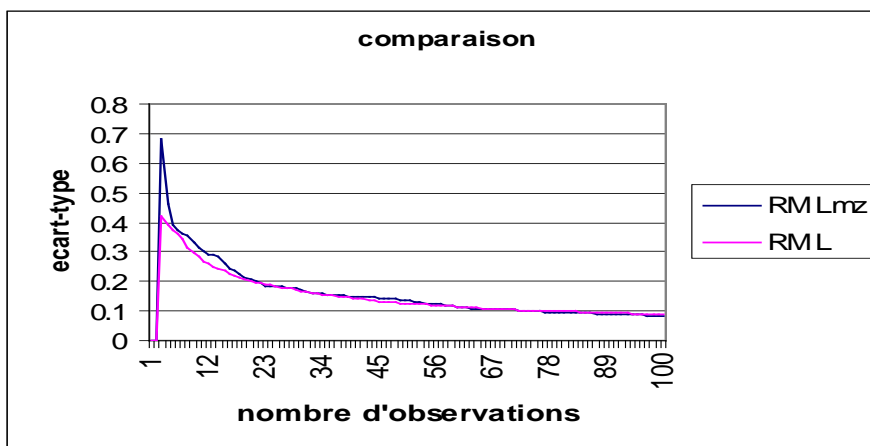
Graphique M_Z101



Graphique ET_Z101

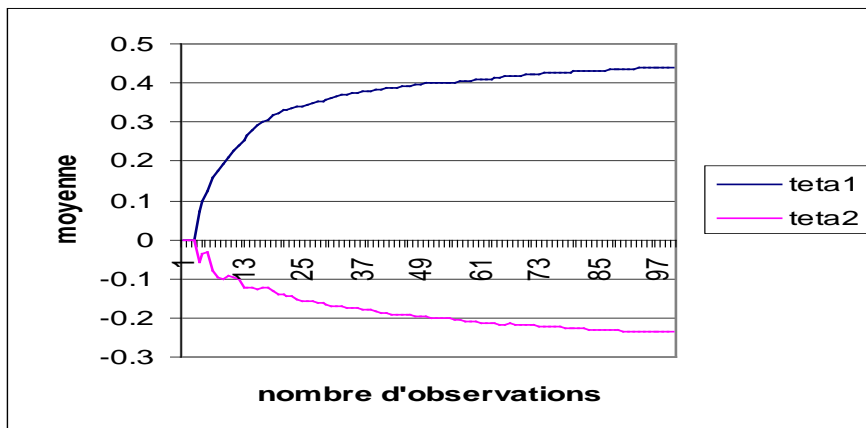


Graphique C_M101

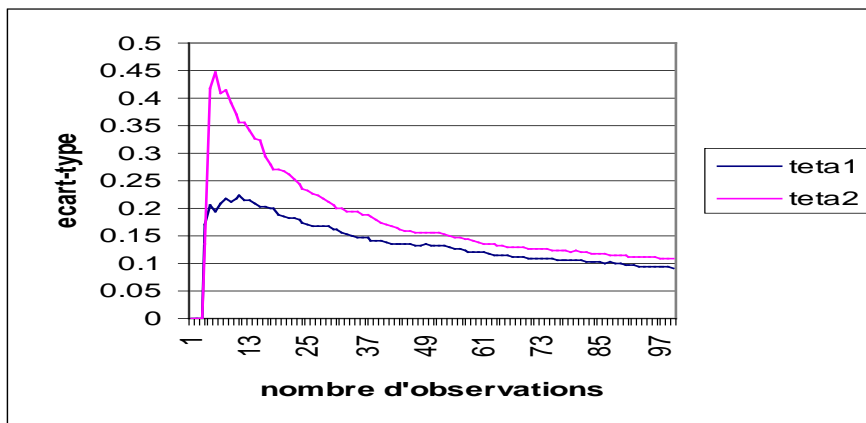


Graphique $C_{ET}101$

Modèle ARMA(0,2)

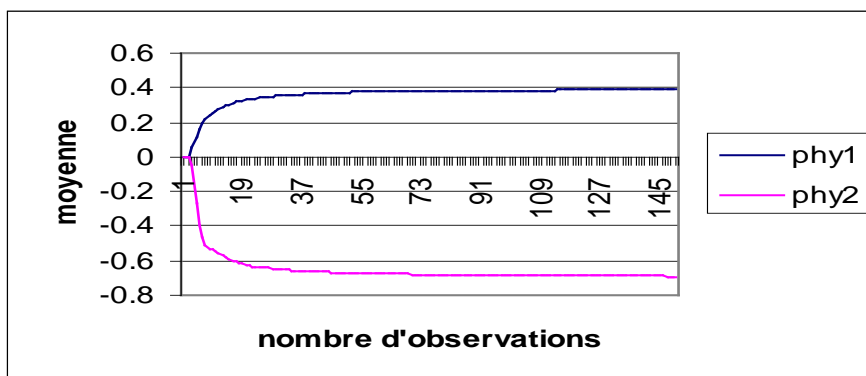


Graphique M_M021

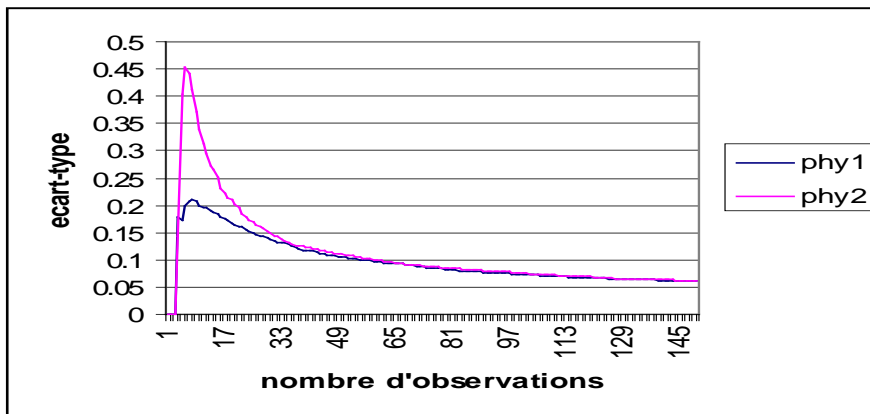


Graphique ET_M021

Modèle ARMA(2,0)

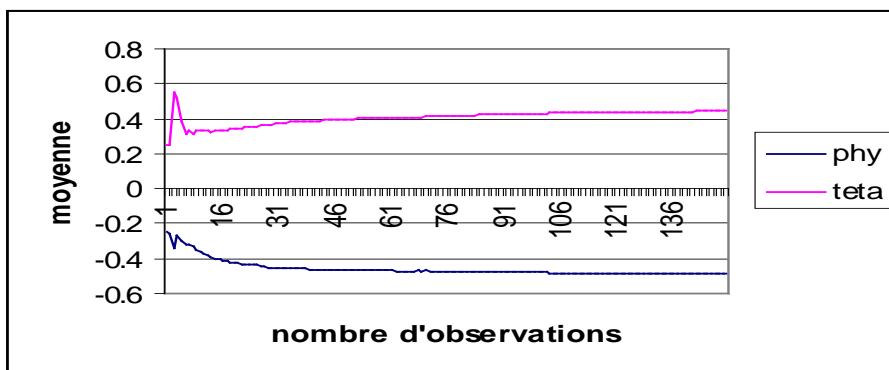


Graphique M_M201

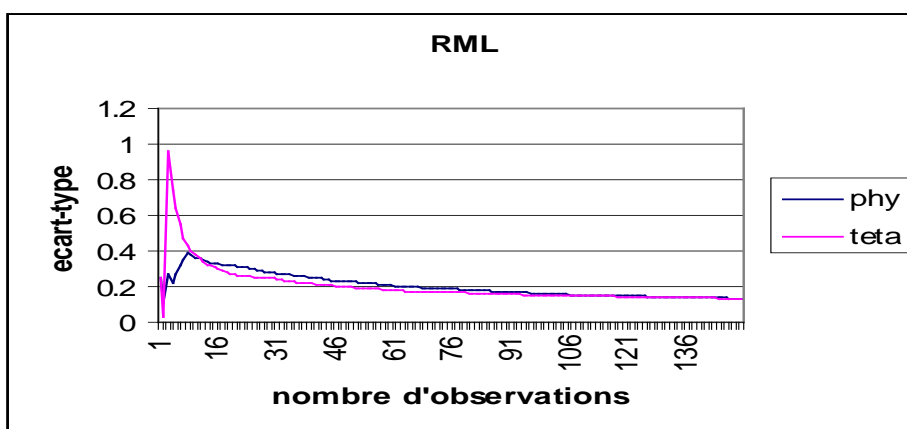


Graphique ET_M201

Modèle ARMA(1,1)

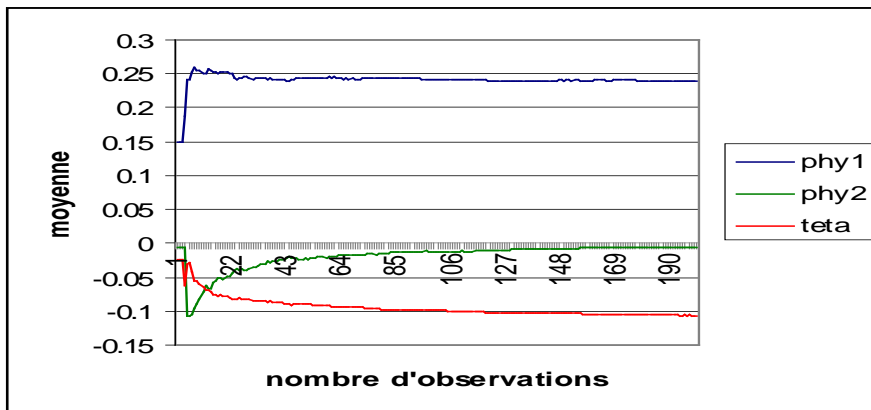


Graphique M_M111

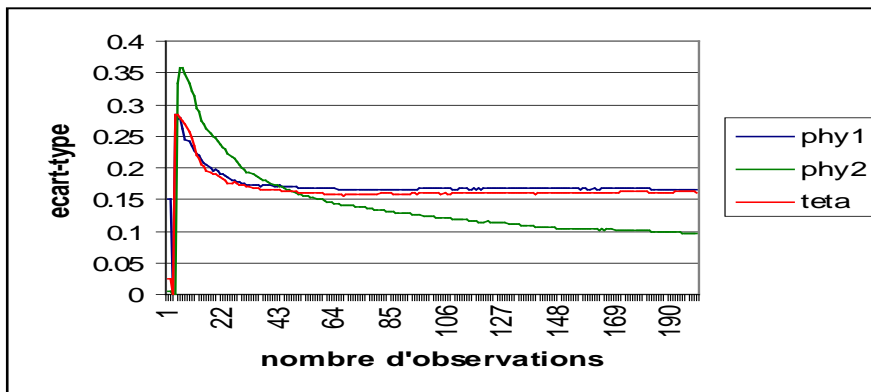


Graphique ET_M111

Modèle ARMA(2,1)

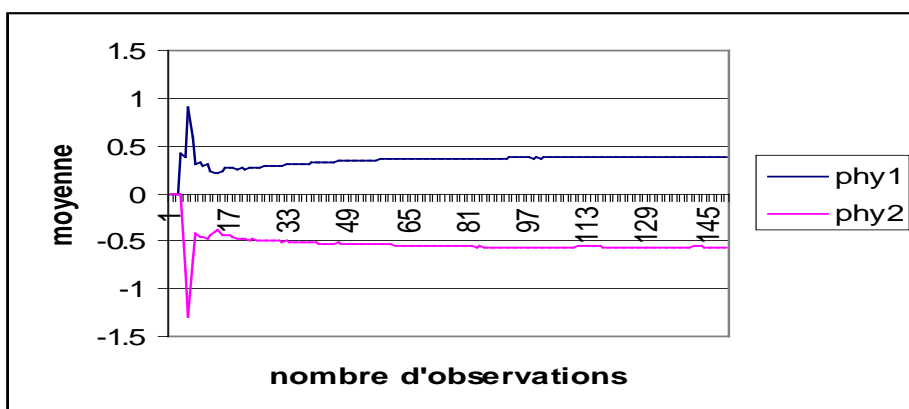


Graphique M_{M211}

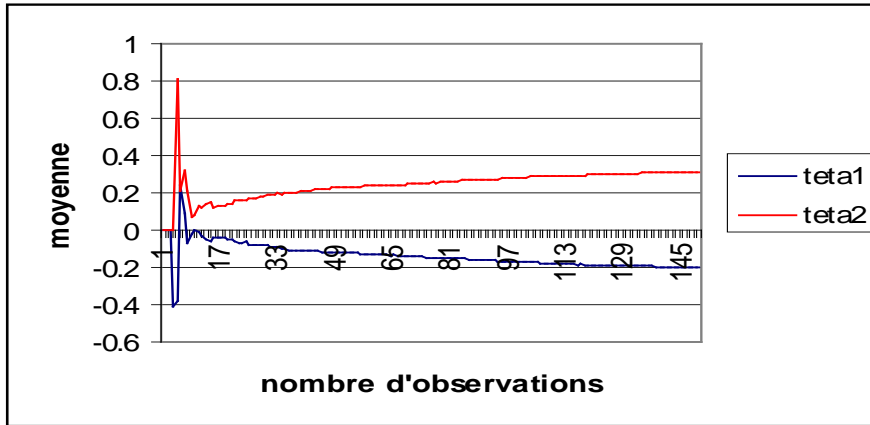


Graphique ET_{M211}

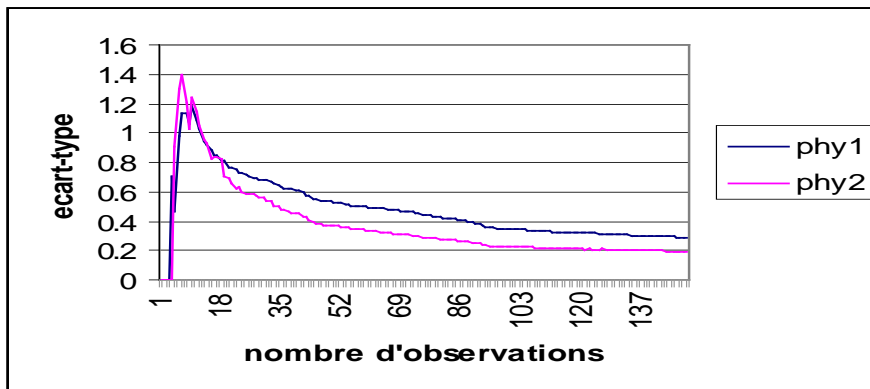
Modèle ARMA(2,2)



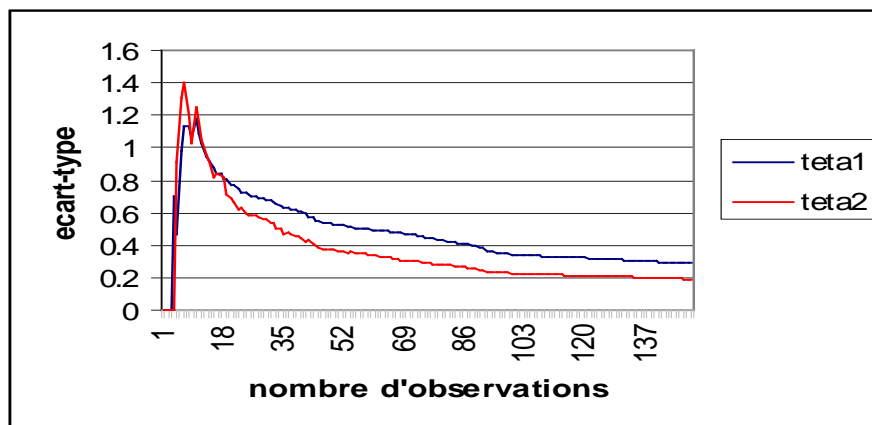
Graphique $M_{M221.1}$



Graphique $M_{M221.2}$

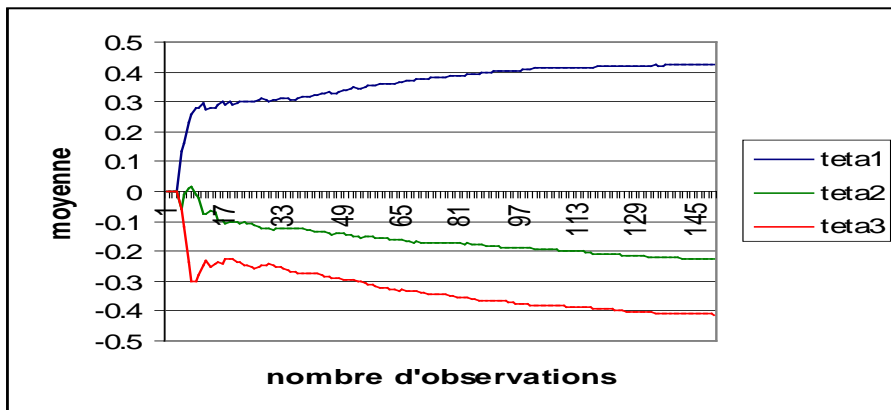


Graphique $ET_{M221.1}$

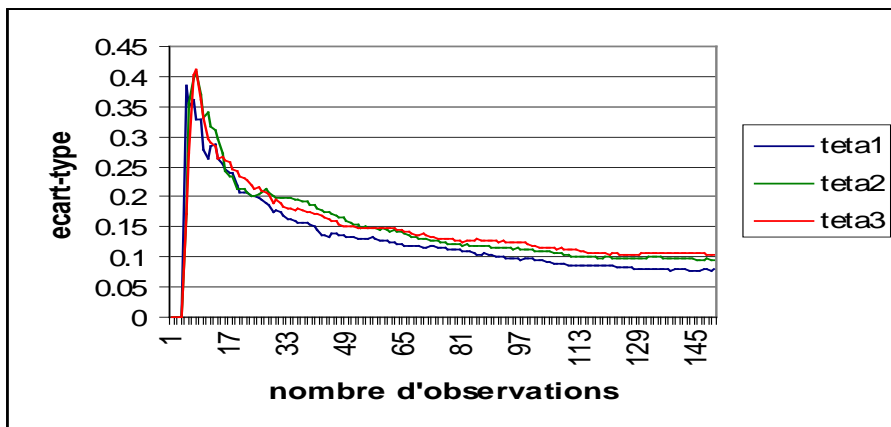


Graphique $ET_{M221.2}$

Modèle ARMA(0,3)

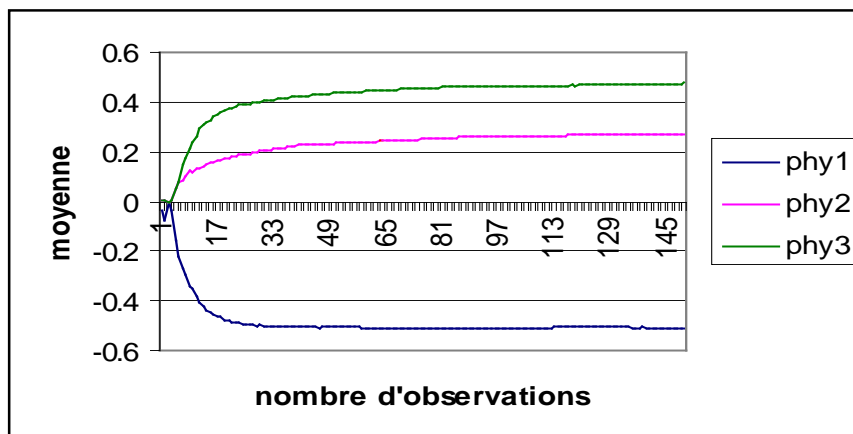


Graphique M_{M031}

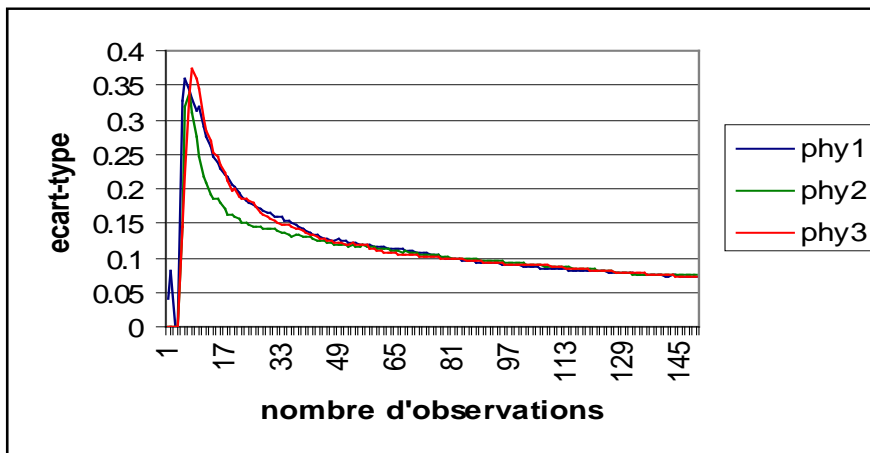


Graphique ET_{M031}

Modèle ARMA(3,0)

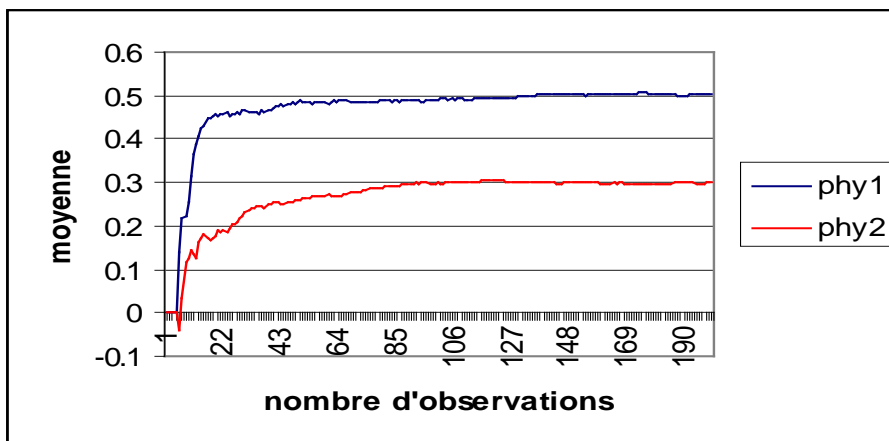


Graphique M_{M301}

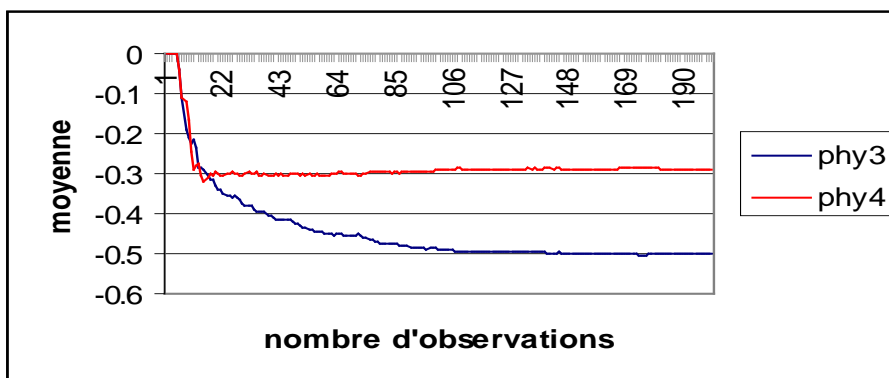


Graphique ET_M301

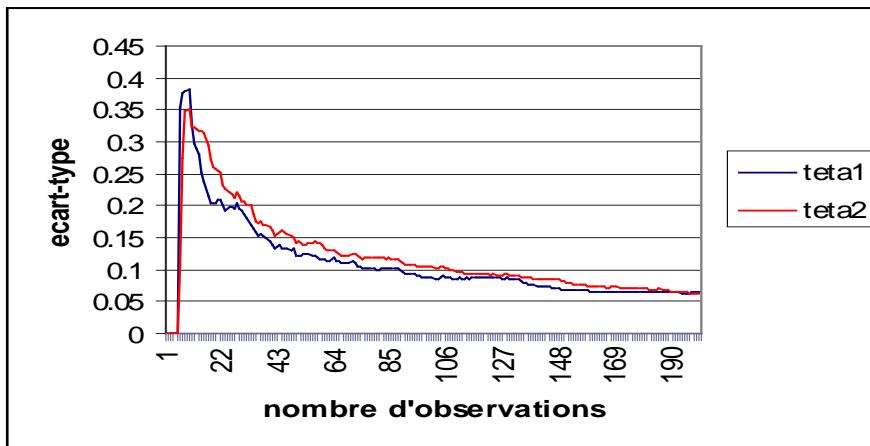
Modèle ARMA(4,0)



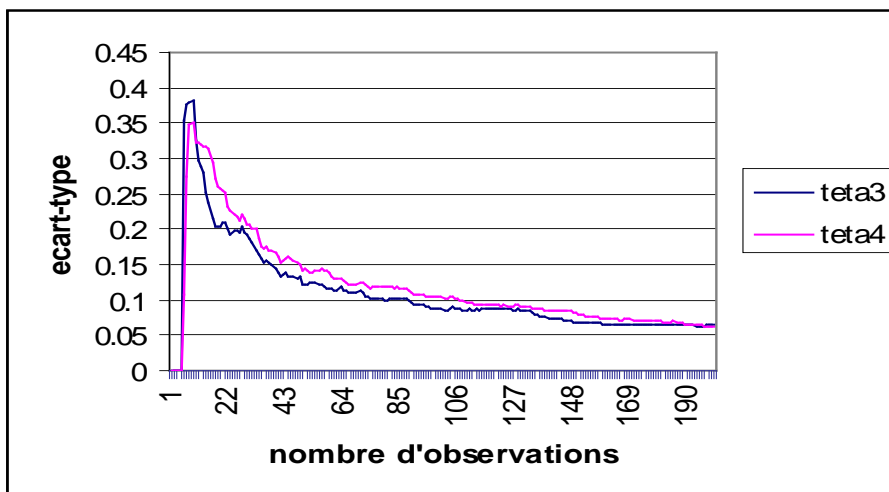
Graphique M_M401.1



Graphique M_M401.2



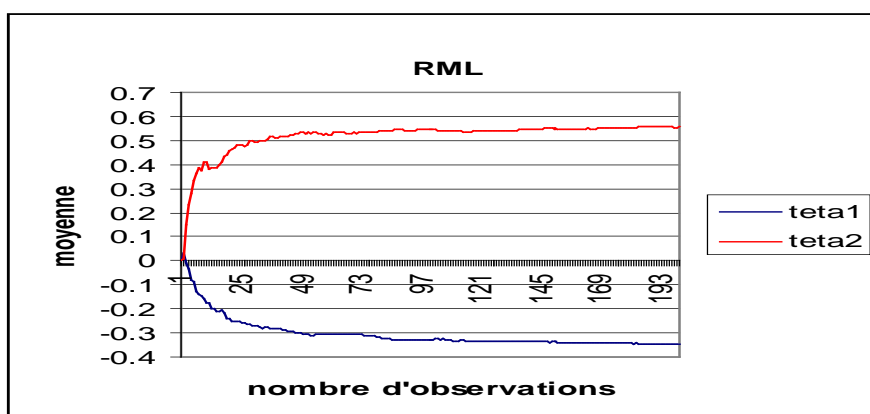
Graphique ET_M401.1



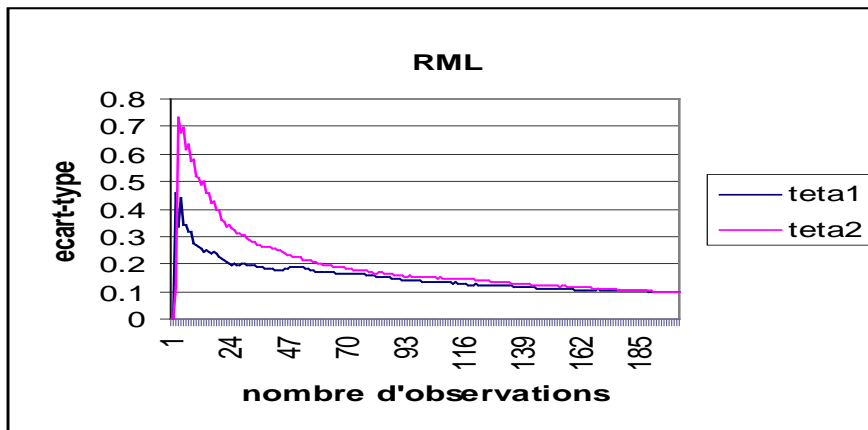
Graphique ET_M401.2

Résultats concernant le cas de modèles ARMA périodiques

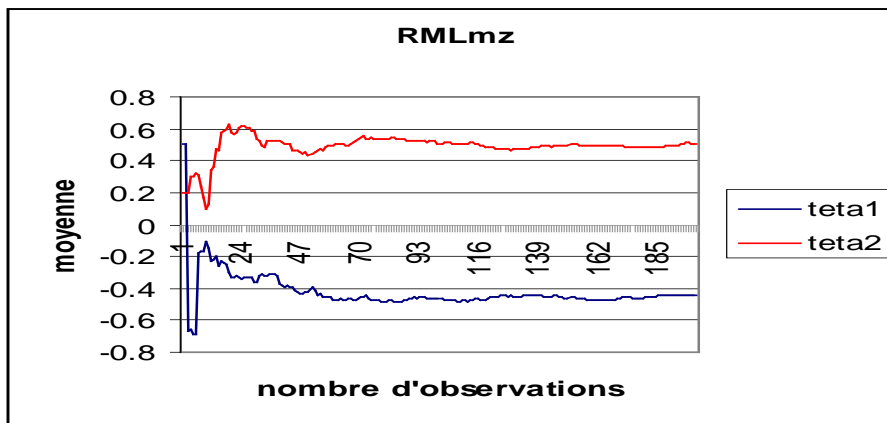
Modèle PARMA₂(0,1)



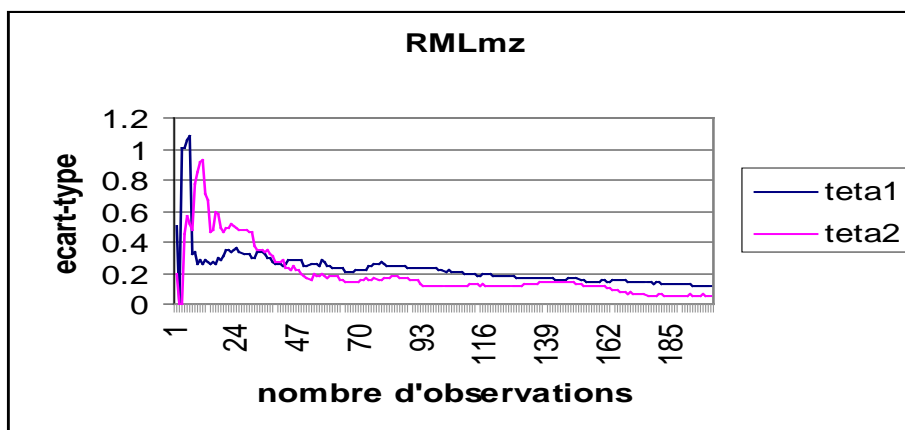
Graphique M_M012



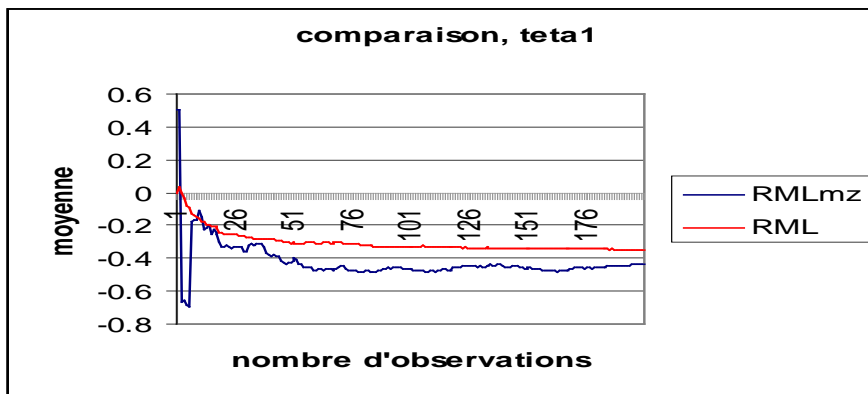
Graphique ET_M012



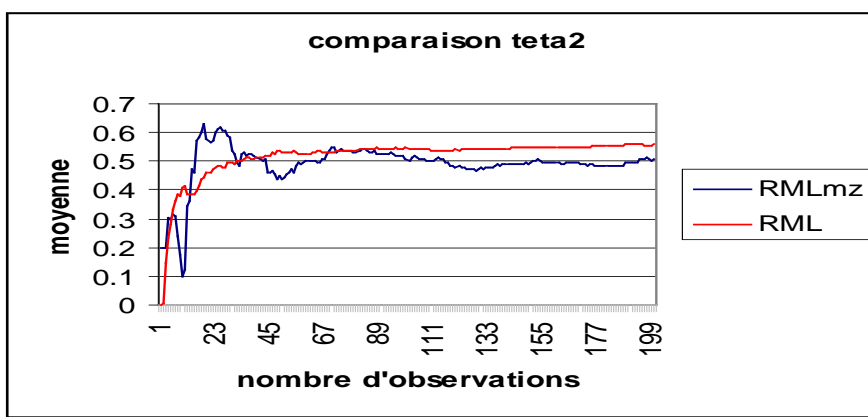
Graphique M_Z012



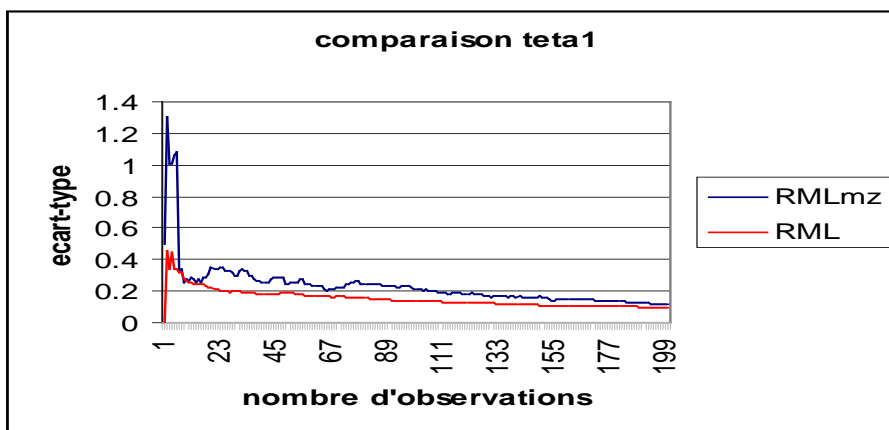
Graphique ET_Z012



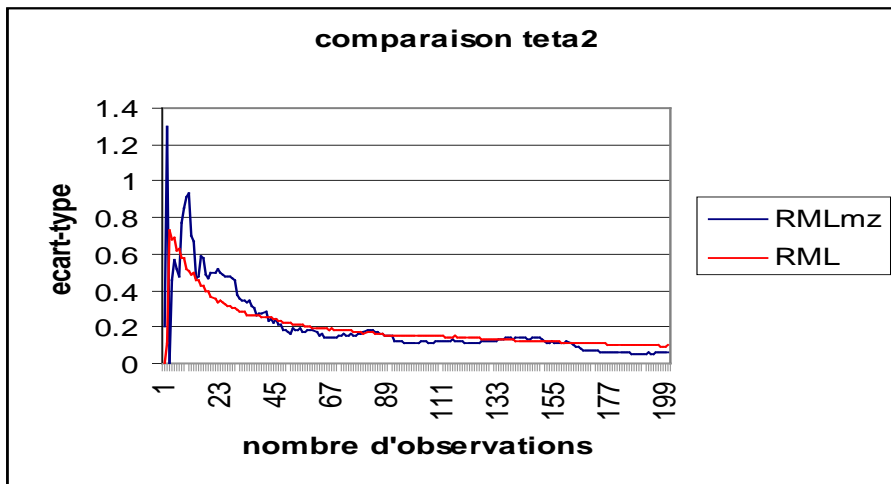
Graphique C_M012.1



Graphique C_M012.2

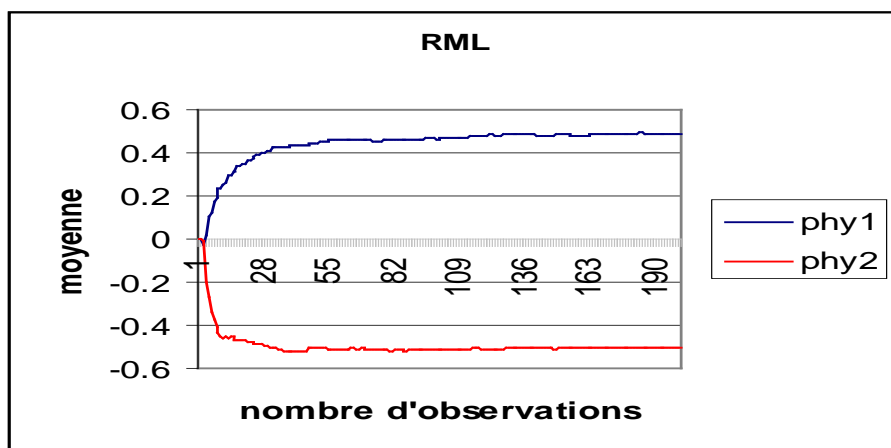


Graphique C_{ET}012.1

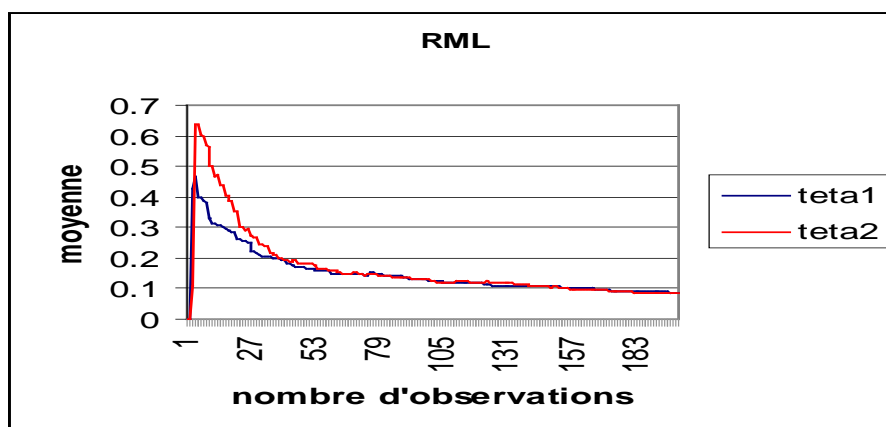


Graphique C_{ET}012.2

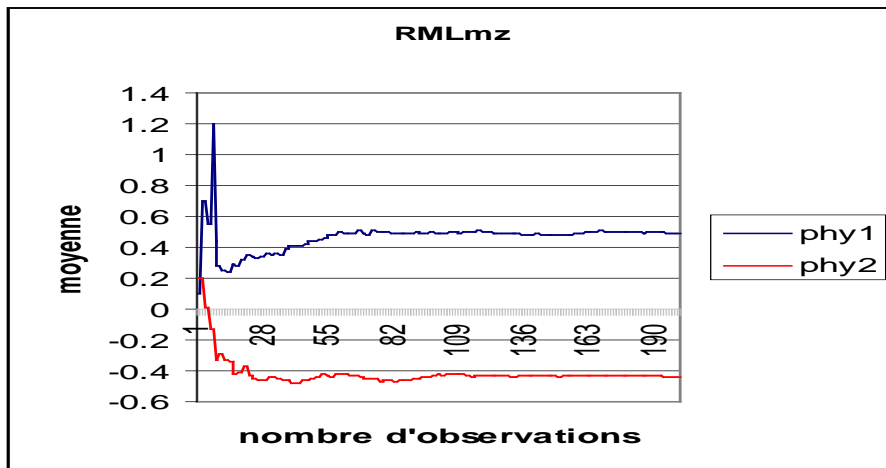
Modèle PARMA₂(1,0)



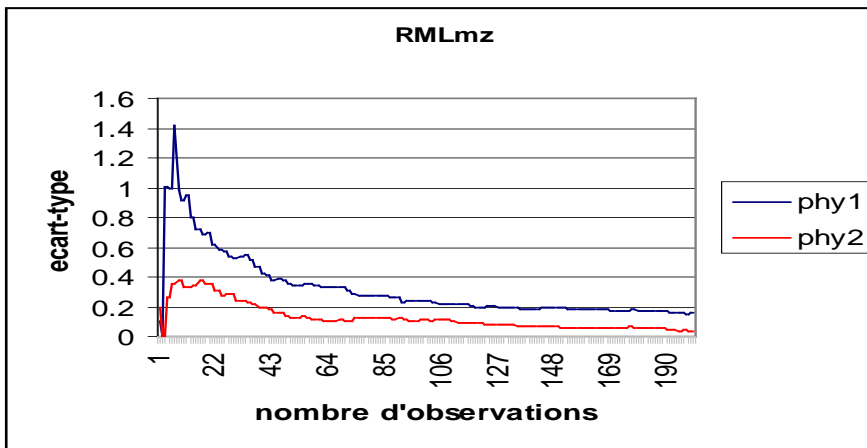
Graphique M_M102



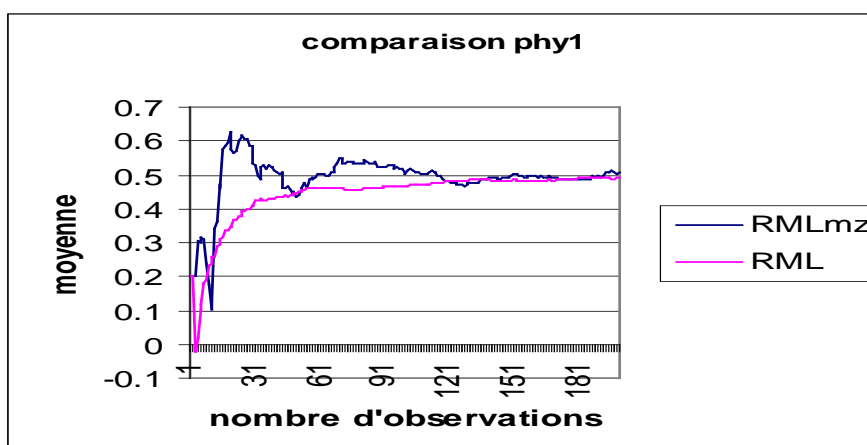
Graphique ET_M102



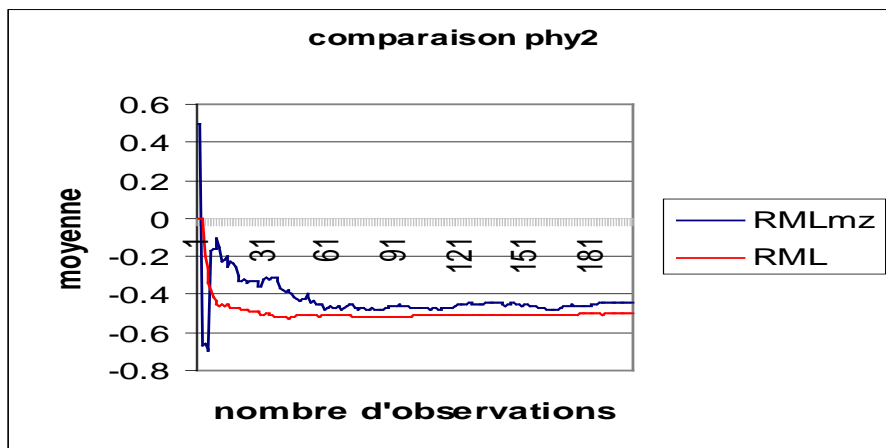
Graphique M_z102



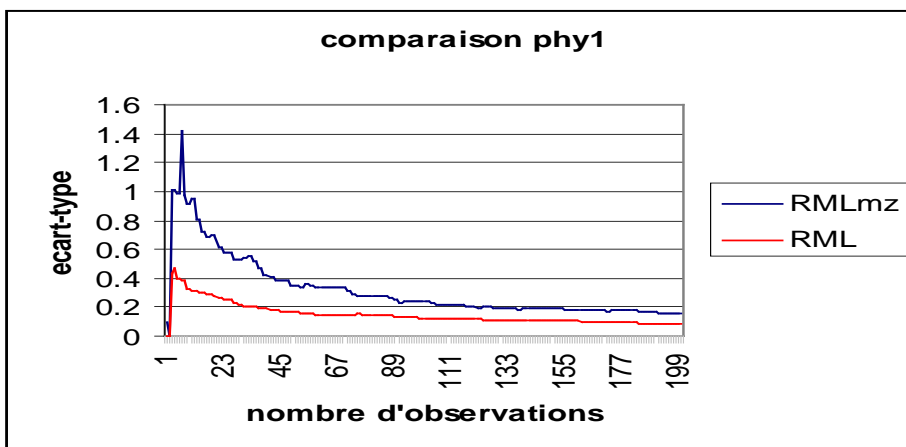
Graphique ET_z102



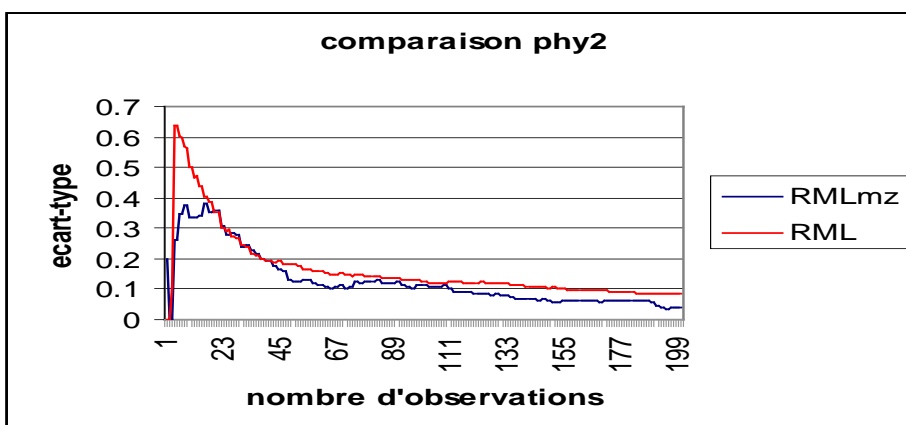
Graphique $C_M102.1$



Graphique C_M102.2

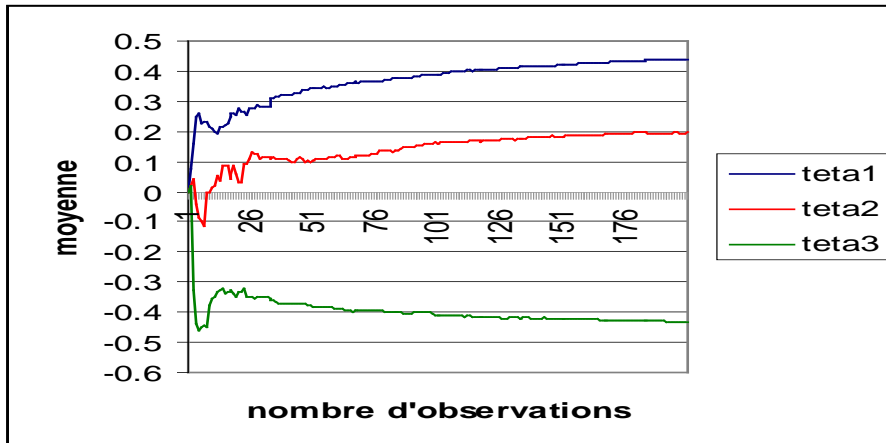


Graphique C_{ET}102.1

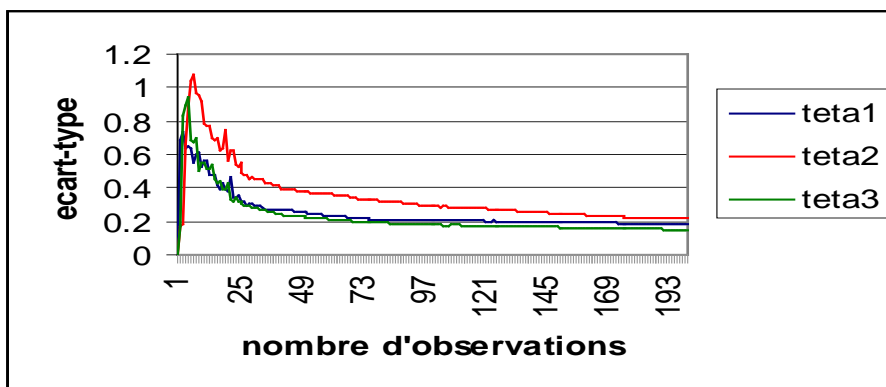


Graphique C_{ET}102.2

Modèle PARMA₃(0,1)

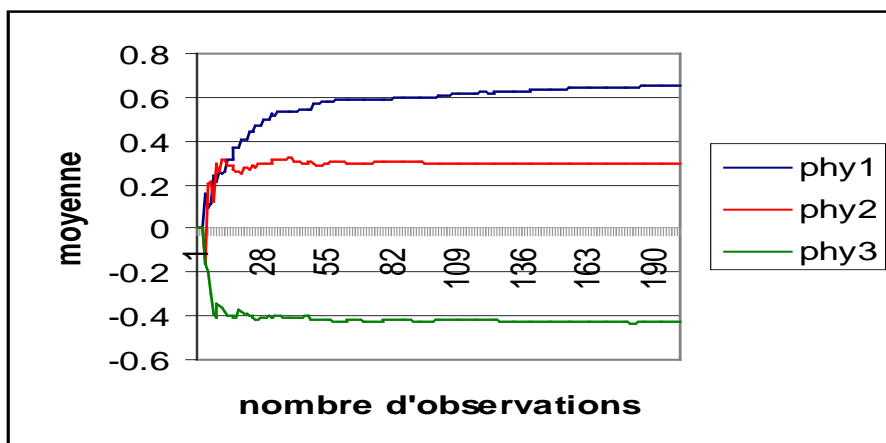


Graphique M_M013

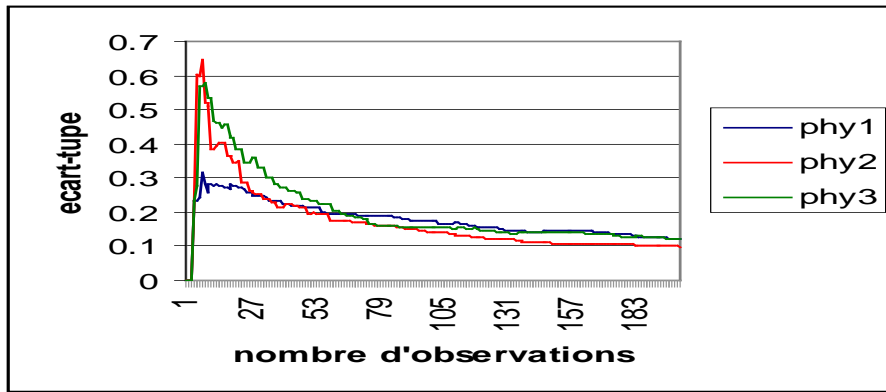


Graphique ET_M013

Modèle PARMA₃(1,0)

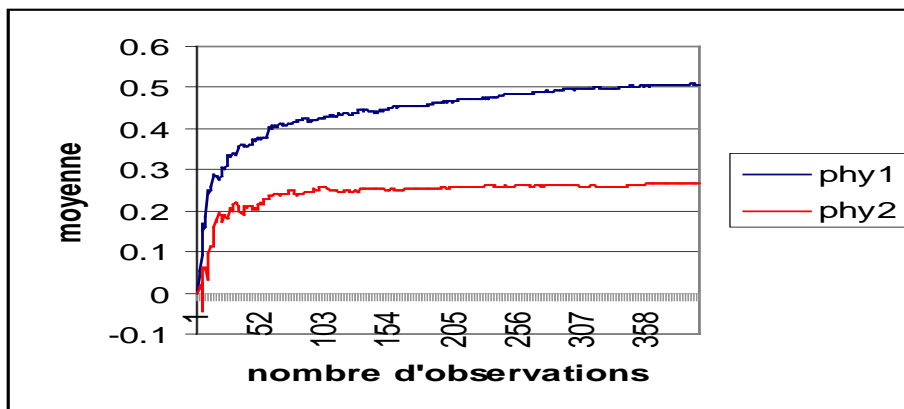


Graphique M_M103

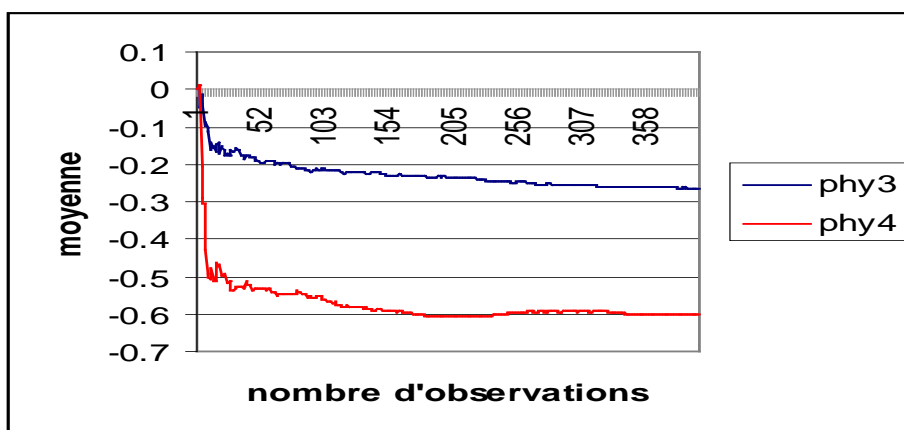


Graphique ET_M103

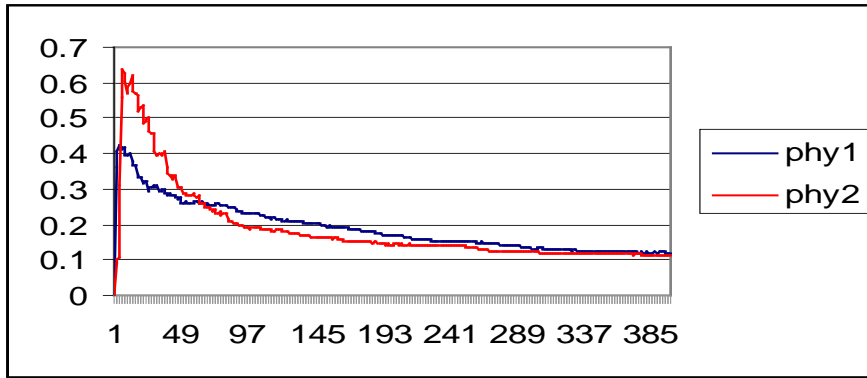
Modèle PARMA₄(0,1)



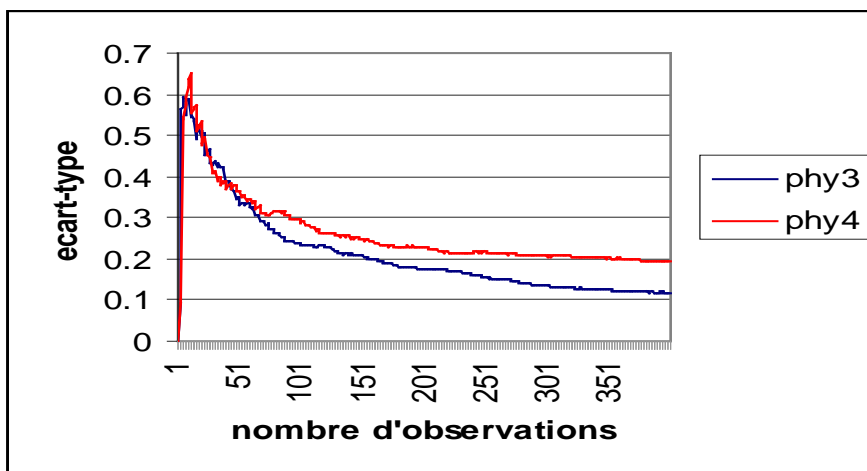
Graphique M_M014.1



Graphique M_M014.2

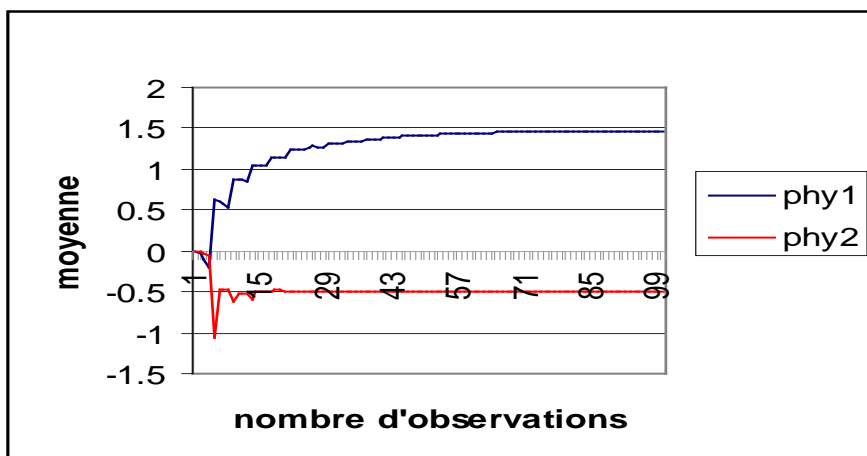


Graphique ET_M014.1

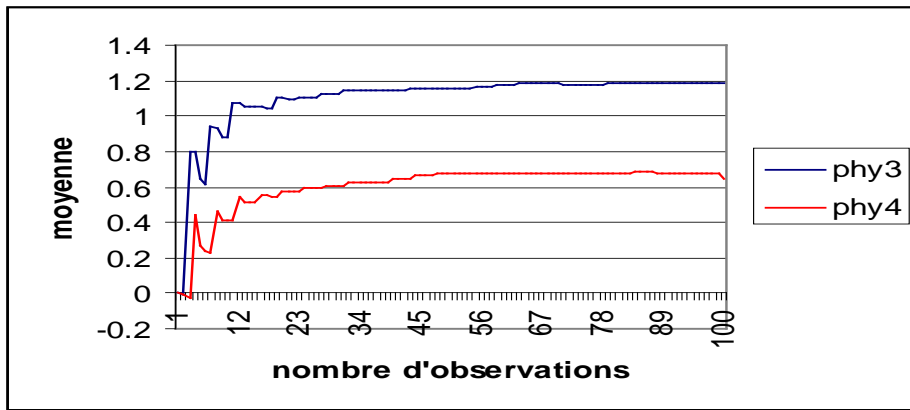


Graphique ET_M014.2

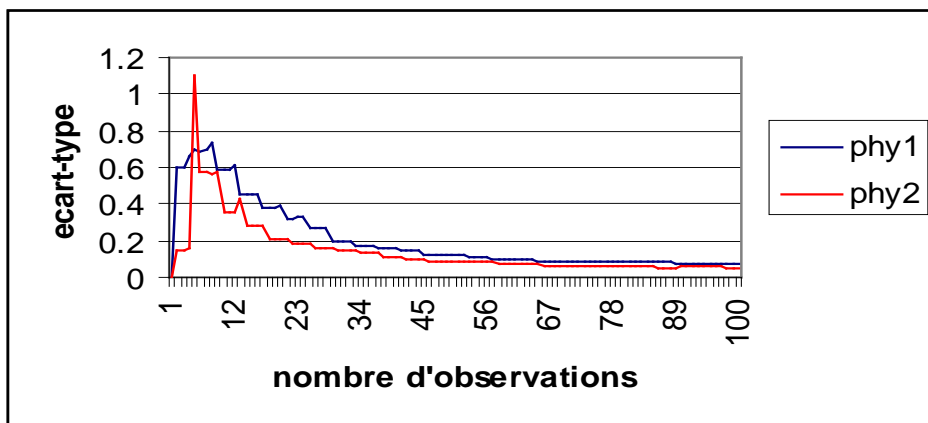
Modèle PARMA₄(1,0)



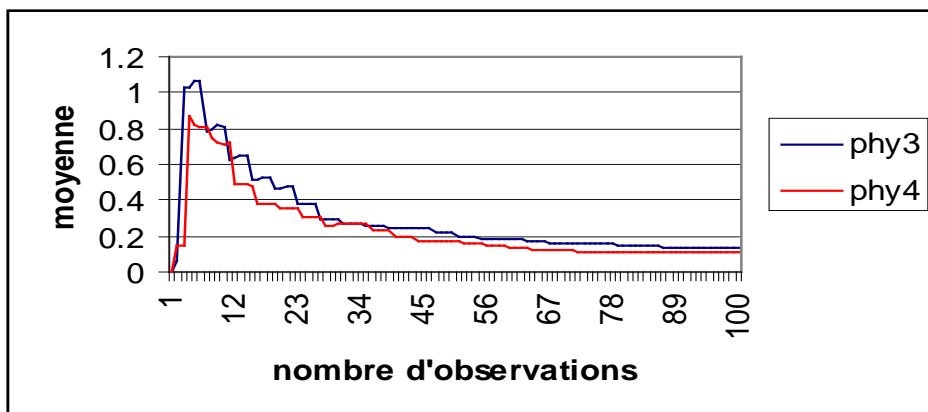
Graphique M_M104.1



Graphique M_M104.2

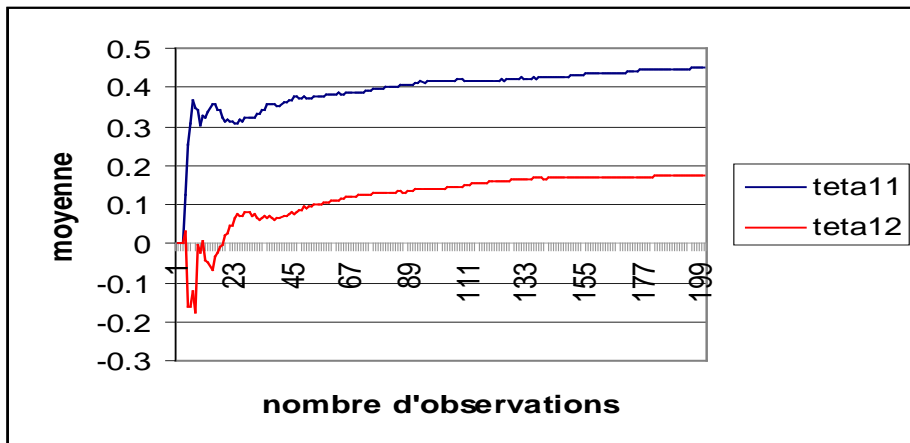


Graphique ET_M104.1

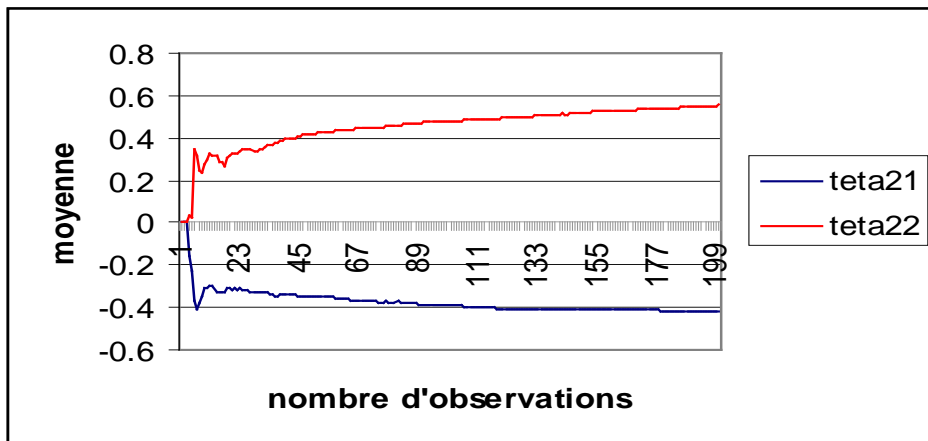


Graphique ET_M104.2

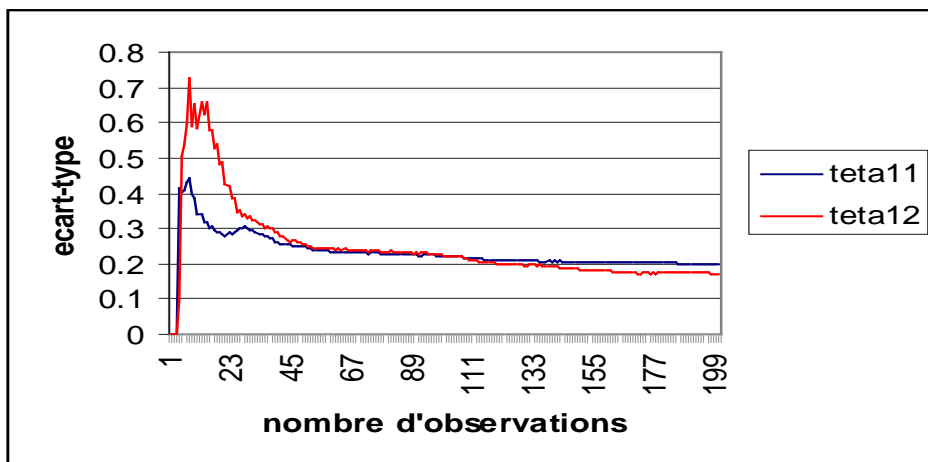
Modèle PARMA₂(0,2)



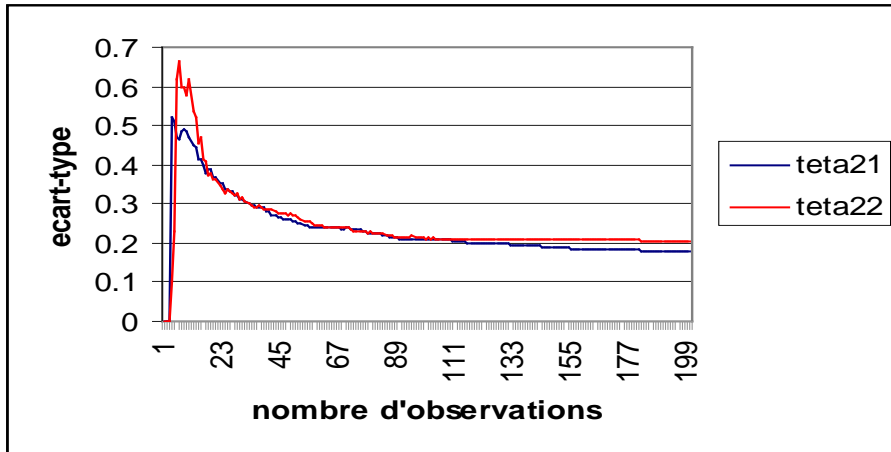
Graphique M_M022.1



Graphique M_M022.2

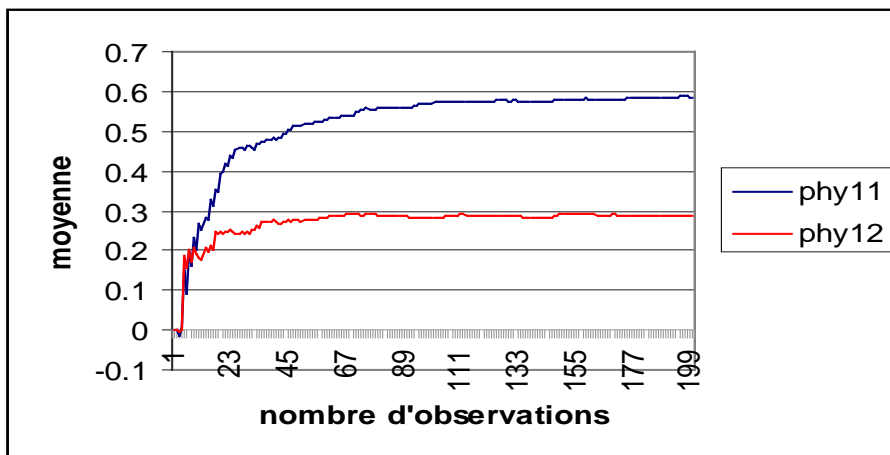


Graphique ET_M022.1

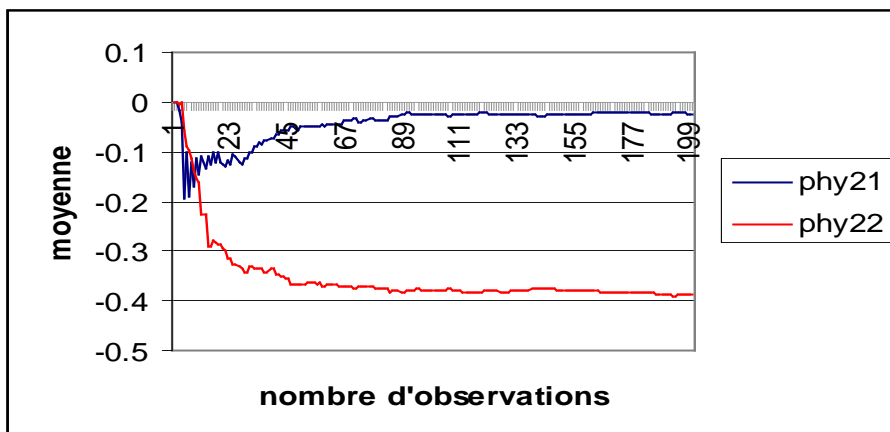


Graphique ET_M022.2

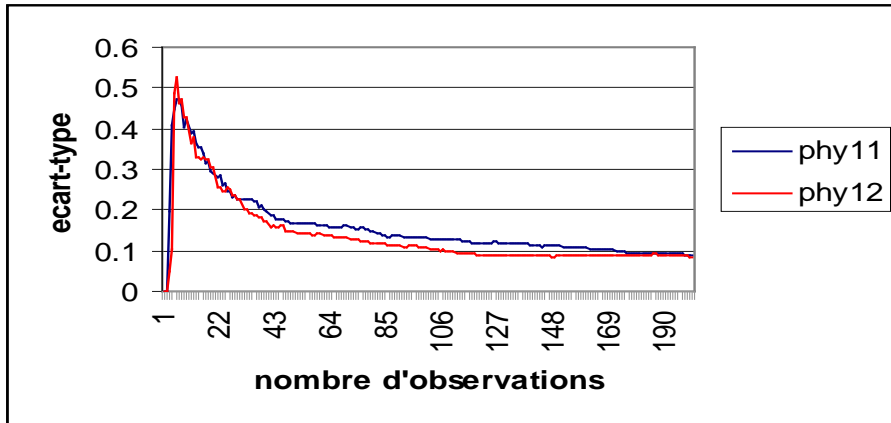
Modèle PARMA₂(2,0)



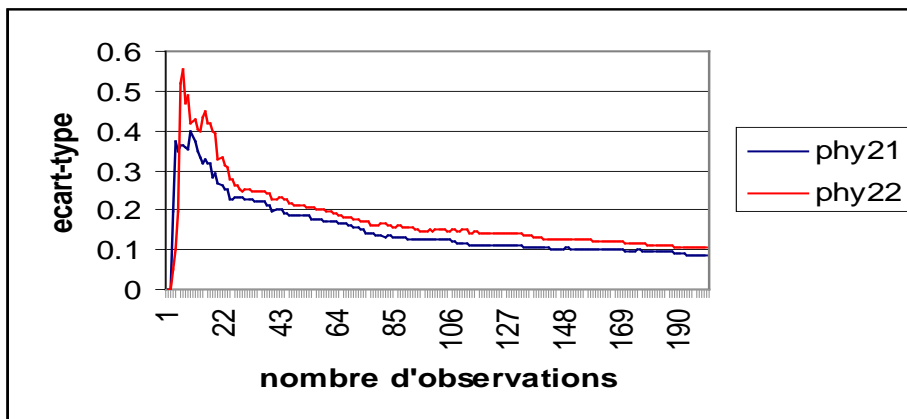
Graphique M_M202.1



Graphique M_M202.2

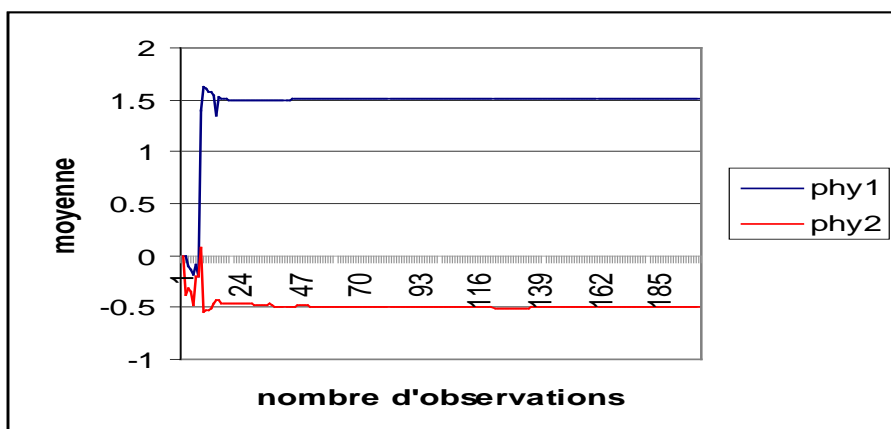


Graphique ET_M202.1

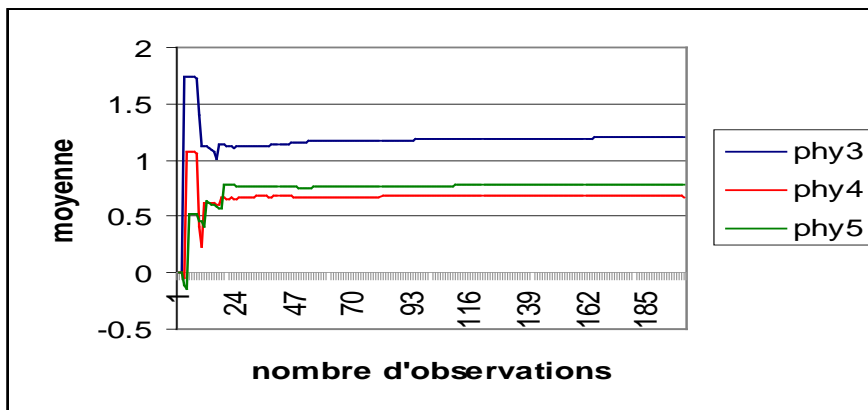


Graphique ET_M202.2

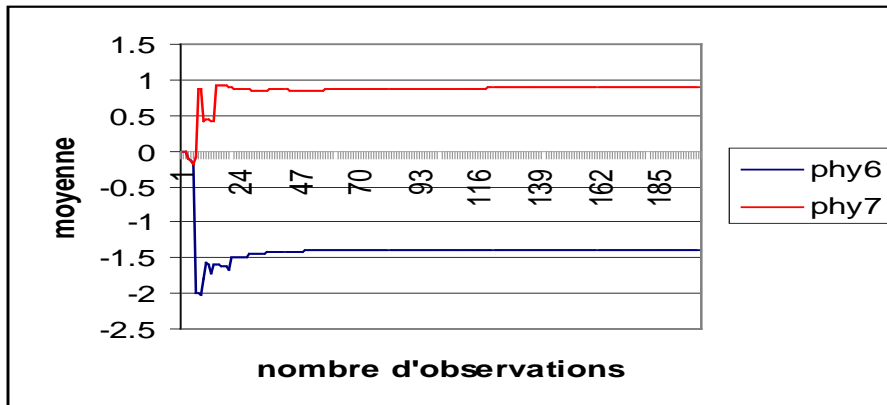
Modèle PARMA₇(1,0)



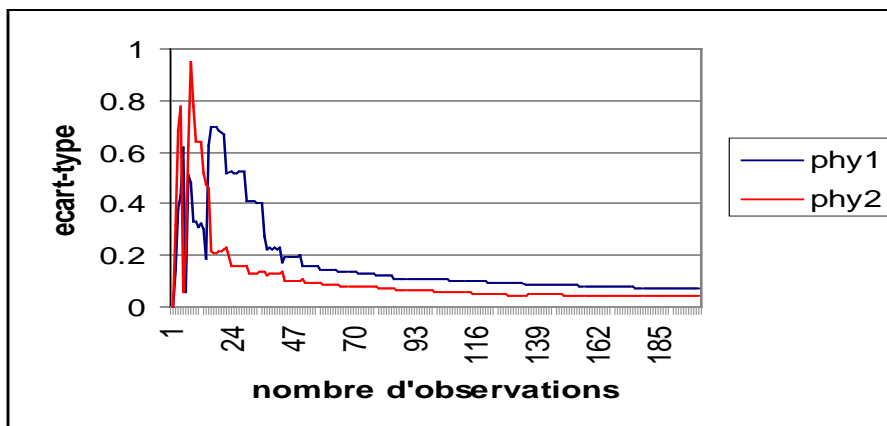
Graphique M_M107.1



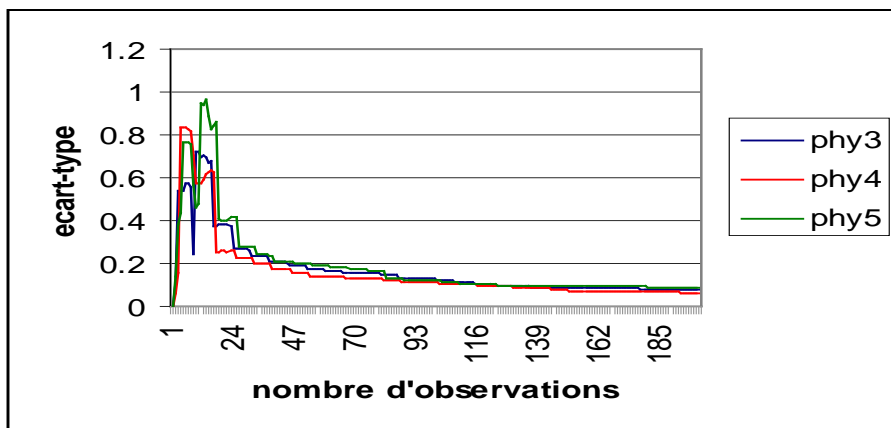
Graphique M_M107.2



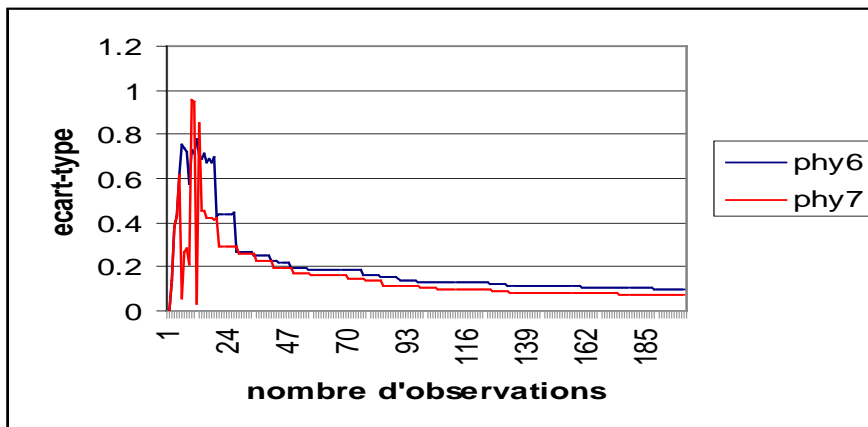
Graphique M_M107.3



Graphique ETm107.1

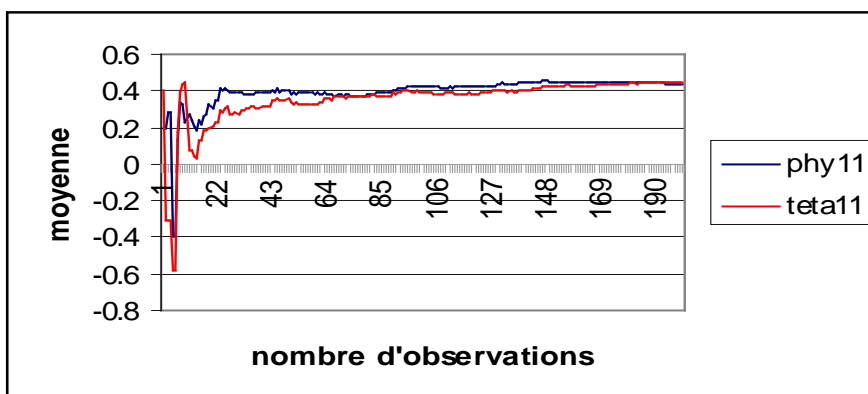


Graphique ETm107.2

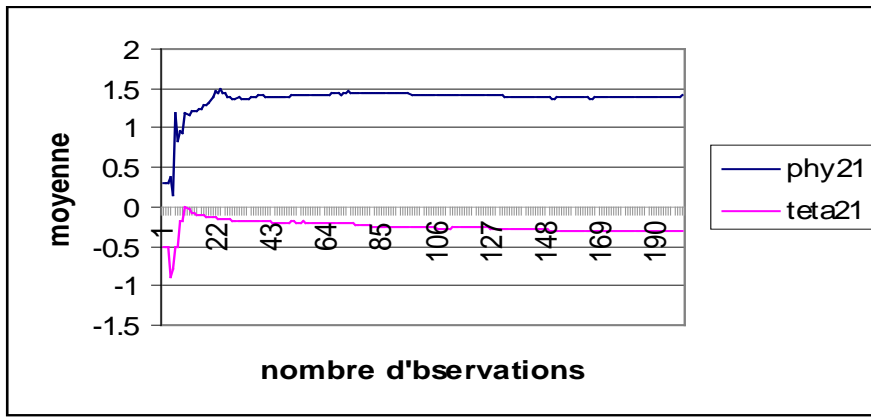


Graphique ETm107.3

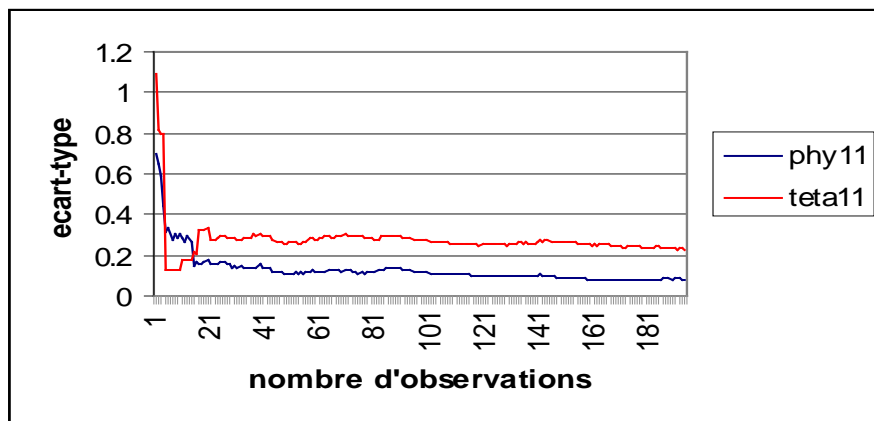
Modèle PARMA2(1,1)



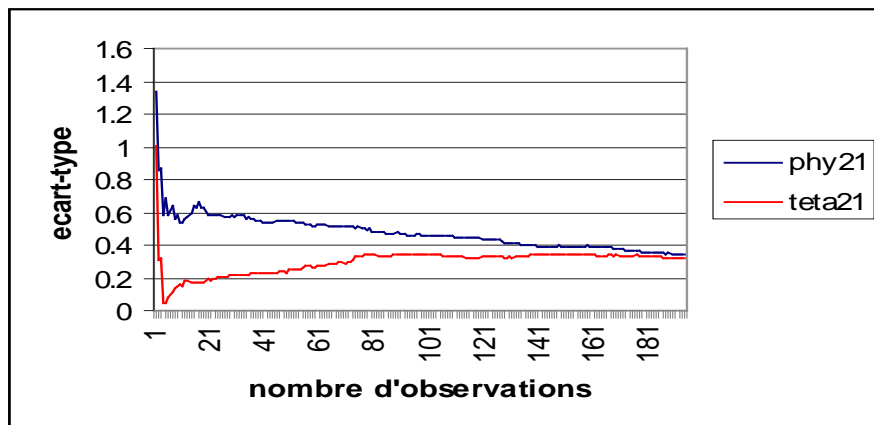
Graphique M_M112.1



Graphique $M_{M112.2}$



Graphique $ET_{M112.1}$



Graphique $ET_{M112.2}$

Conclusion

Ce travail a été consacré à l'étude du problème de l'estimation des paramètres de modèles ARMA périodiques (PARMA) selon deux contextes différents : estimation en présence de données hors-ligne et de données en-ligne. Notons que nous avons travaillé dans la majorité des cas sous l'hypothèse de la normalité des erreurs. Cette hypothèse semblant être restrictive peut être complétée dans le cas général par l'utilisation de quelques transformations adéquates, citons en particulier la transformation de Box-Cox.

Dans la première partie de ce travail, nous nous sommes intéressés à l'estimation hors-ligne des modèles PARMA, nous avons présenté une sorte de synthèse (mais non exhaustive) de l'ensemble des méthodes d'estimation disponibles. Nous avons donc distingué deux grandes classes de méthodes ; une première classe utilisant directement les données, à savoir la méthode des moindres carrés et ses variantes, et la méthode du maximum de vraisemblance et ses approximations. Et une seconde classe qui utilise des transformations des données en résumés statistiques, consistant à résoudre certaines équations basées sur ces résumés. Nous citons en particulier la méthode de Yule-Walker périodique (Pagano (1978)) d'une part et son amélioration du point de vue algorithmique dans le cas autorégressif périodique d'autres part ; appelée la méthode récursive de Durbin-Levinson périodique (Sakai (1982)). Il est à noter que dans le cas autorégressif moyenne mobile périodique la méthode de Boshnakov (1996) se substituera à celle de Sakai (1982). Du point de vue asymptotique, les deux classes de méthodes ont des propriétés statistiques équivalentes dans le cas autorégressif périodique, et l'utilisation des méthodes récursives s'avèrent la plus intéressante compte tenu de la simple complexité algorithmique. Mais dans le cas de l'existence de composantes moyennes mobiles, la méthode du maximum de vraisemblance semble être la meilleur, pas du point de vue asymptotique seulement mais aussi du point de vue algorithmique. Ces résultats bien connus dans le cas de modèle ARMA classique ont été confirmés dans la classe de modèles ARMA périodiques à travers une étude de simulation comparative sur différents modèles choisis.

La seconde partie de ce travail a été consacré à l'estimation des modèles PARMA en présence des données en ligne. Nous avons considéré à ce sujet la classe de méthodes issues des méthodes d'estimation hors-ligne, plus particulièrement la méthode de l'erreur de prévision minimale (RPEM). Pour cette même classe de méthodes nous nous sommes intéressés à la conception et à la formulation de quelques algorithmes disponibles dans le cas ARMA classique. Nous avons tout d'abord, présenté l'algorithme des moindres carrés récursifs (RLS) appliqué aux modèles PAR, discuté les détails de son application puis abordé ses propriétés statistiques. Par ailleurs, nous avons généralisé l'algorithme RML (Recursive Maximum Likelihood) concernant les modèles ARMA classiques au cas d'un modèle ARMA périodique. Cet algorithme, qui peut être considéré comme généralisation de l'algorithme RLS au cas de l'ajout d'une composante moyenne mobile a été appliqué sur plusieurs instances avec succès. Nous avons ensuite généralisé l'algorithme RML_{MZ} , amélioration due à Zahaf (1998) de l'algorithme RML dans le cas ARMA classique, aux cas de modèles PARMA. La stabilité numérique et la précision des résultats ont été également vérifiées dans le cas de la généralisation de cet algorithme. Notons que l'algorithme RML_{MZ} fait intervenir à chaque itération le calcul de la matrice d'information de Fisher d'un modèle PARMA, faute d'indisponibilité d'algorithmes précis pour cette classe de modèles, nous avons donc introduit un algorithme à cette fin en généralisant, aux cas de modèles périodiques, l'algorithme de Klein et Mélard (1989). Nous avons aussi présenté une procédure pour le calcul des autocovariances d'un modèle PAR qui se trouve être une tâche nécessaire pour l'élaboration de l'algorithme de calcul de la matrice d'information de Fisher. Les algorithmes récurrents présentés dans la deuxième partie de notre travail se basent sur des routines d'optimisation stochastiques à cause de la complexité de la fonction critère. Comme ces procédures sont souvent locales, ce qui renforce la dépendance de l'algorithme récurrent à quelques conditions expérimentales ; diverses précautions seront à prendre. Aussi, nous avons consacré tout un chapitre mettant en relief les différents aspects pratiques de l'implémentation numérique. En fin, nous avons vérifié les performances des algorithmes présentés par le biais d'une étude de simulation intensive.

Nous avons vu que par analogie de la méthodologie de Box-Jenkins concernant la construction, basée sur des données hors ligne, de modèles ARMA ; la méthodologie de l'identification en-ligne introduite par les ingénieurs dans le contexte de données en ligne se réduit quant à elle, à trois étapes essentielles, en outre l'identification du modèle paramétrique, la formulation et la conception de l'algorithme récurrent à travers l'optimisation stochastique et l'analyse asymptotique des propriétés statistiques de ce dernier. Concentré essentiellement sur la deuxième étape, la deuxième partie de notre travail se complétera d'une part, par l'analyse des propriétés statistiques des algorithmes RML et RML_{MZ} périodiques ; d'autres part la littérature sur l'étape de l'identification du modèle qui est équivalente à celle de Box-Jenkins est assez large et abondante.

Malgré que les estimateurs récurrents aient des qualités voisines des estimateurs du maximum de vraisemblance, les méthodes récurrentes d'estimation même accompagnées de la simple complexité algorithmique n'ont pas encore connu autant de succès dans le domaine de la statistique, en effet la littérature de leur application est relativement maigre par rapport à celle des ingénieurs. Cependant, avec la complexité croissante du champ d'application et avec l'augmentation constante des performances des calculateurs ces méthodes connaîtront sans doute un grand intérêt.

Références

- Abdrabo, N.A and Priestly, M.B. (1967). On the prediction of nonstationary processes, *Journal of the Royal Statistical Society, Series B.* **29**, pp 570-585.
- Adams, G. J. and Goodwin, G. C (1995). Parameter estimation for periodic ARMA Models. *Journal of Time Series Analysis.* **16**, pp 127-145.
- Akaike, H. (1973). Block Toeplitz matrix inversion, *SIAM Journal of Applied Mathematics*, **24**, pp 234-241.
- Ali, M. M. (1977). Analysis of autoregressive-moving average processes, *Annals of Institute of Statistics and Mathematics*, **26**, pp 363-387.
- Andel, J. (1983). Statistical analysis of periodic autoregression. *Aplikace Matimatiky*, pp 364-385.
- Anderson, O. D. (1975). The recursive nature of the stationary and inversibility remains on the parameters of mixed autoregressive-moving average process, *Biometrika* **62**, pp 704-706.
- Anderson, P. L. and Vecchia, A. V. (1993). Asymptotic results for periodic autoregressive moving-average process. *Journal of Time Series Analysis.* **14**, pp 1-18.
- Astrom, K. J. (1972). Unpublished notes. In Ljung and Söderström (1983).
- Astrom, K. J. (1980). Maximum likelihood and prediction error methods, *Automatica*, **16**, pp 457-476.
- Astrom, K. J. and Bohlin, T. (1965). Numerical identification of linear dynamic systems from normal operating record, *IFAC symposium on Self-Adaptive Systems*, Teddington, England.
- Benouamer, Z. and Bentarzi, M. (1996). Identification des ordres d'un modèle ARMA périodique : méthode du coin. *International conference on applied mathematics and engineering science. CIMAI' 1996.* Casablanca-MAROC.
- Bentarzi, M. and Hallin, M. (1993). Autoregression periodique ou autoregrssion stationnaire? Tests localement asymptotiquement optimaux. *XXVèmes journées de statistiques*, Vannes, France 24-28 Mai 1993.
- Bentarzi, M. and Hallin, M. (1994). On the Inversibility of Periodic Moving Average Models *Journal of Time Series Analysis.* **15**, pp 263-268.

- Bentarzi, M. (1995). "Modèles de séries chronologiques à coefficients périodiques". *Thèse de Doctorat es sciences*. Institut de mathématiques, U.S.T.H.B, Alger, Algérie.
- Bentarzi, M. (1995). On the Wold-Cramér decomposition of periodically correlated m-variate q-dependent processes. (à paraître).
- Bentarzi, M. (1998). Model Building Problem of periodically m-variate moving average process. *Journal of Multivariate Analysis*. **66**, pp 1-21.
- Bentarzi, M. (2001). Predictive density order selection of periodic autoregressive models. A paraître.
- Benveniste, A. et Crespel, P.(1981). Algorithmes adaptatifs et approximations stochastiques, *outils et modèles mathématique pour l'automatique, l'analyse des systèmes et le traitement du signal*, v. **1**, CNRS, Paris, pp 593-630.
- Benveniste, A. Métivier, M. et Priouret, P.(1990). Adaptive algorithms and stochastic approximation, *Springer-Verlag, Berlin*.
- Bhuiya, R. K. (1971). Stochastic analysis of periodic hydrologic processes. *Journal of Hydraulic Division of the American Society of Civil Engineering*. **97**, pp 949-962.
- Bierman, G. J. (1977). Factorization methods for discrete sequential estimation, Academic Press, New York.
- Boshnakov, G. N. (1994a). A generalization of the Burg's algorithm to periodically correlated time series. À paraître.
- Boshnakov, G.N. (1994b). Periodically correlated sequences: some properties and recursion. Research Report 1, Division of Quality Technology and Statistics, Luelo University, Sweden.
- Boshnakov, G. N. (1996). Recursive computation of the parameters of periodic Autoregressive moving-average processes. *Journal of Time Series Analysis*. **17**, pp 333-349.
- Box, G.E.P. et Jenkins, G. M.(1976). Time Series Analysis, Forecasting and Control, Holden-Day, (revised edition), San Francisco, CA.
- Brockwell, P. J. and Davis, R. A (1988). Time Series : Theory and methods. Springer-Verlag. New York.
- Burg, J. P. (1975). Maximum entropy spectral analysis, Ph. D. dissertation, Department of Geophysics, Stanford University, Stanford, CA.
- Cipra, T. (1985). Periodic moving average processes. *Aplikace matematiky* **30**, pp 218-229.

- Cleveland, W.P et Tiao, G.C. (1979). Modeling seasonal time series. *Revue Economique Appliquée* **32**, pp 107-129.
- Cramér, H. (1961). On some classes of nonstationary stochastic processes. *Proceeding of the 4th Berkley Symposium on Mathematical Statistics and Probability*. **2**, pp 57-78.
- Dent, W. (1977). Computation of the exact likelihood function of an ARIMA process, *Journal of Statistical Computational Simulation*, **5**, pp 193-206.
- Djuric, P. M. and Kay, M. (1992). Order selection of autoregressive models. *IEEE Transactions on signal processing*, **42**, pp 2829-2833.
- Duffin, E. J. (1969). Algorithms for classical stability problems, *S.I.A.M. Revue*. **11**, pp 196-213.
- Duflo, M. (1990). Méthodes récursives aléatoires. *Techniques stochastiques. Massons*. Paris.
- Durbin, J. (1960). The fitting of time series models, *Revue Internationale of Institut of Statistics* **28**, pp 233.
- Favier, G. (1982). Filtarage, Modélisation et identification de systèmes linéaires stochastiques à temps discret. Edition du C.N.R.S. Paris.
- Franke, J. (1985). A Levinson-Durbin recursion for autoregressive moving average process. *Biometrika* **72**, pp 573-581.
- Franses, P. H. (1993). A multivariate approach to modeling univariate seasonal time series. *Journal of Econometrics* **63**, pp 133-151.
- Franses, P. H. (1996). Periodicity and Stochastic Trends In Economic *Times Series*. Oxford University Press.
- Gardner, G. Harvey, A. C. et Philips, G. D. A. (1980). Algorithm AS154. An algorithm for exact maximum likelihood estimation of autoregressive-moving average models by means of Kalaman filtering, *Journal of the Royal Society Series C. Applied Statistics.*, **29**, pp 311-322
- Ghysel, E. and Hall, A. (1992a). Testing periodicity in some linear macroeconomic models. *Discussion Paper*, C.R.D.E., Université de Montréal.
- Ghysel, E. and Hall, A. (1992b). Lagrange multiplier test for periodic structure. *Unpublished manuscript Paper*, C.R.D.E., Université de Montreal.
- Gladyshev, E. G. (1961). Periodically correlated random sequences. *Soviet Matematika*. **2**, pp 385-88.
- Hall, A. D. and Nicholls, D. F. (1980). The evaluation of exact maximum likelihood estimates for VARMA models, *Journal of Statistical Computational and Simulation*, **10**, pp 251-262.

- Hallin, M.(1984). Spectral factorization of nonstationary moving average processes. *The Annals of Statistics*, **12, 1**, pp 172-192.
- Hallin, M.(1986). Nonstationary q-dependent processes and time-varying moving average models : invertibility properties and the forecasting problem. *Advanced Applied Probability* **18**, pp 170-210.
- Hannan, E. J. (1970). *Multiple Time Series*, John Wiley, New York.
- Hanan, E. J. (1980). Recursive estimation based on ARMA models, *Annals of Statistics*, **8**, pp 762-777.
- Harti, M. (1996). "Algorithmes pour l'estimation par pseudo-maximum de vraisemblance exacte pour des modèles VARMA sous forme classique et sous forme structurée", Thèse de doctorat, Institut de statistique, Université Libre de Bruxelles.
- Hemis, R. (1999). "Sélection de l'ordre des modèles autorégressifs périodiques, Approche classique et approche bayésienne". Thèse de Magistère. Institut de mathématiques, U.S.T.H.B, Alger, Algérie.
- Hilmer, S. C., and Tiao, G. C. (1979). Likelihood Function of Stationary Multiple Autoregressive Moving Average Models, *Journal of the American Statistical Association*, **74**, pp 652-660.
- Jiménez, C., MacLeod, A. I. and Hipel, K. W. (1989) Kalaman filter estimation for periodic autoregressive moving average models. *Stochastic Hydrologic and Hydraulic* **3**,pp 227-40.
- Jones, R.H et Brelsford, W. M. (1967). Time series with periodic structure. *Biometrika* **54**, pp 403-408.
- Kalaman, E. (1960). A new approach to linear filtering and predicting problems, *Transactions of the ASME, journal of Basic Engineering Series*, **82D**, pp 35-45.
- Kennedy, W.H. and Gentle, J.E. (1980). *Statistical computing*. Marcel Decker, New York.
- Kiehm, J. L and Mélard, G. (1981). ARIMA models with time-dependent coefficients for economic time series. *In Time Series Analysis* pp 355-368. *O. D. Anderson and M. R. Perryman, eds. Amsterdam, North -Holland.*
- Klein, A. and Mélard, G. (1989). On algorithms for computing the covariance matrix of estimates in autoregressive-moving average models. *Computational Statistics Quarterly*, **5**, pp 1-9.

- Klein, A. and Mélard, G. (1990). Fisher's information matrix for seasonal autoregressive-moving average models, *Journal of Time Series Analysis*, **11**, pp 231-237.
- Klein, A. and Mélard, G. (1994a). Computation of the Fisher information matrix for SISO models, *IEEE Trans. Signal Processing*, **42**, pp 684-688.
- Klein, A. and Mélard, G. (1994b). The information matrix of multiple input single output time series models, *J. Computational and applied mathematics*, **51**, pp 349-356
- Klein, A. and Mélard, G. and Zahaf, T. (1998). Computation of the exact information matrix of Gaussian dynamic regression time series models, *Annals of Statistics*, **26**, à paraître.
- Kushner, H. J. et Clark, D. S. (1978). Stochastic approximation methods for constrained and unconstrained systems, *Applied Mathematics Science. Series, 26*, Springer Verlag, Berlin.
- Li, W. K and Hui, Y. V (1988). An Algorithm for the exact likelihood of periodic autoregressive moving average models. *Communication of Statistics and Simulation*, pp 1483-1494.
- Ljung, G. M. et Box, G. E. P. (1979). The likelihood function of stationary autoregressive-moving average models, *Journal of the Royal Society Series C Applied. Statistics*, **24**, pp 255-259
- Ljung, L. (1977). Analysis of recursive stochastic algorithms, *IEEE Transactions on Automatic Control*, AC **22**, pp 551-575.
- Ljung, L. (1978). Convergence analysis of parametric identification methods, *IEEE Transactions on Automatic Control*, AC **23**, pp 770-783.
- Ljung, L. and Soderstrom, T. (1983). Theory and practice of recursive identification, MIT Press, Cambridge, Massachusetts.
- Lund, R. and Basawa, I. V. (1998) Recursive Prediction and likelihood evaluation for periodic ARMA models. A paraître.
- McLeod, A. I. (1975). The Derivation of the theoretical autocovariances function of autoregressive-moving average time series, *Applied Statistics* **24**, pp 255-256.
- McLeod, A. I. (1992a). Diagnostic checking of periodic autoregression models with application. *Journal of Time Series Analysis* **15**, pp 221-233.
- McLeod, A. I. (1992b). Parsimony, model adequacy and periodic correlation in time series forecasting. *International statistical review*. **61**, pp 387-393.
- Mélard, G. (1984a). Algorithm AS197: A fast algorithm for the exact likelihood of autoregressive-moving average models, *Journal of the Royal Statistics Society Series C Applied Statistics*, **33**, pp 104-114.

- Mélard (1984b). Analyse des séries chronologiques, Journées d'étude en Statistiques, Publication de l' A. S. U.
- Mélard, G. (1985a). Exact derivatives of the likelihood of ARMA processes. In *Proceedings of the Statistical Computing Section*, American Statistical Association, Washington D. c., **pp** 187-192.
- Mélard, G. (1985b). *Analyse de Données chronologiques*, Séminaire de Mathématiques supérieurs, Séminaire scientifique OTAN (NATO Advanced Study Institute), Presses de l'université de Montréal, Montréal.
- Mélard, G. (1989). Méthode d'identification récurrente pour la modélisation des données chronologiques économiques, Actes du Salon Int. de l'Inf. de la Tél., séance *III*: Intelligence artificielle et génie logiciel, Association Marocaine des utilisateurs de l'informatique, Direction de la statistique, Rabat.
- Mélard, G. et Klein, A. (1994). On a fast algorithm for the exact information matrix of a Gaussian ARMA time series, *IEEE Trans. Signal Processing*, **42**, **pp** 2201-2203.
- Michaut, F. (1992). Méthodes adaptatives pour le signal. Hermès, Paris.
- Moore, J. B. et Ledwich, G. (1980). Multivariable adaptive parameter and state estimators with convergence analysis, *Journal of Australian Mathematical Society*, **B, 21**, **pp** 176-197.
- Morf, M., Sidhu, G. S. et Kailath, T. (1974). Some new algorithms for recursive estimation on constant, linear, discrete-time systems, *IEEE Transaction on Automatics and Control*, **19**, **pp** 315-323.
- Newbold, P. (1974). The exact likelihood function for mixed autoregressive-moving average process, *Biometrika*, **61**, **pp** 423-426.
- Newton, H. J. (1978). The information matrices of the parameters of multiple mixed time series, *Journal of Multivariate analysis*, **8**, **pp** 317-323.
- Newton, H.J. (1982). Using periodic autoregressions for multiple spectral estimation. *Technometrics*, **6**, **pp** 1310-1317.
- Osborn, D. R. (1992). The implication of periodically varying coefficients for seasonal time-series processes. *Journal of Econometrics* **48**, **pp** 373-384.
- Osborn, D. R. and Smith, J.P. (1989). The performance of periodic autoregressive models in forecasting seasonal U.K. consumption. *Journal of Business and Economic Statistics* **7**, **pp** 117-127.
- Pagano, M. (1978). On periodic and multiple autoregression. *Annals of statistics*, **6**, **pp** 1310-1317.

- Parzen, E. et Pagano, M. (1979). An approach to modeling seasonally stationary time series, *Journal of Econometrics*, **9**, pp 137-153. North-Holland Publishing Company.
- Pearlman, J. G. (1980). An algorithm for the exact *likelihood of a high-order autoregressive-moving average process*, *Biometrika*, **67**, pp 232-233.
- Peiris, M. S. (1984a). Nonstationary first order autoregression: some multivariate extension. *Statistics Research Report Paper. 92*. Department of Mathematics. Monash University. Australia.
- Peiris, M. S. (1984b). Recursive relations and linear prediction for the nonstationary multivariate first order AR models. *Statistics Research Report. Paper 98*. Department of Mathematics. Monash University. Australia.
- Pham, D. T. (1986). An efficient algorithm for maximum likelihood estimation for autoregressive moving average model, *Rapport de recherche, TIM3 RR 613-M*, Informatique et Mthématique Appliquées, Grenoble.
- Porat, B. et Friedlander, B. (1986a). Bounds on accuracy of Gaussian ARMA parameter estimation methods based on sample covariance, *IEEE Trans. On Auto. Cont. AC* **31, 6**, pp 579-582.
- Porat, B. et Friedlander, B. (1986b). Computation of the exact information matrix of Gaussian time series with stationary random components, *IEEE Trans. Acoustic, Speech, Signal Processing*, **14**, pp 118-130.
- Robbins, H. and Monro, S. (1951). A stochastic approximation method. *Annals of Mathematical Statistics*. **22**, pp 702-710.
- Sakai, H. (1982). Circular Lattice Filtering Using Pagano's Method. *IEEE Transactions on Acoustics, Speech, and Signal Processing ASSP-30*, pp 279-286.
- Salas, J.D, Duane C. Boes and Ricardo A. smith. (1982). Estimation of ARMA Models with Seasonal Parameters. *Water Resources Research*, **18**, pp 1006-1010.
- Shaman, P. (1973). On the inverse of the covariance matrix for an autoregressive process, *Biometrika*, **60**, pp 193-196.
- Siddiqui, M. M. (1958). On the inversion of the sample covariance matrix in stationary autoregressive process, *Annals of Mathematical Statistics*, **29**, pp 585-588.
- Singh, N and Peiris, M. S (1987). A note on the properties of some nonstationary ARMA processes. *Stochastic Processes and their Applications*, **24**, pp 151-155.

- Soderstrom, T. (1973). An on-line algorithm for approximate maximum likelihood identification of linear dynamic systems. Report 7308, *Department of Automatics and Control*, Lund Institute of technology. Lund, Sweden.
- Solo, V.(1979). The convergence of AML, *IEEE Transaction On Automatics and Control*, **AC 24**, pp 985-963.
- Solo, V. (1980). Some aspects of recursive parameter estimation, *international Journal of control*, **32**, pp 395-410.
- Solo, V. (1981). The second order properties of a time series recursion, *Annals of Statistics*, **9**, pp 307-317.
- Talbi, D. and Bentarzi, M. (1996). Estimation adaptive d'un modèle autorégressif périodique causal. *International conference on applied Mathematics and engineering sciences. CIMAI' 1996*. Casablanca-MAROC.
- Tiao, G. C. et Ali, M. M. (1971). Analysis of correlated random effects: linear model with two random components, *Biometrika*, **58**, pp 37-51.
- Tiao, G.C et Grupe, M.R. (1980). Hidden periodic autoregressive-moving average models in time series data. *Biometrika* **67**, pp 365-373.
- Terceiro Lomba, J. (1990). Estimation of Dynamic Econometric Models with Errors in Variables, Springer-Verlag, Berlin.
- Troutman, B.M. (1979). Some results in periodic autoregression. *Biometrika*, **66**, pp 219-237.
- Tsay, R.S. and Tiao, G.C. (1984). Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models. *Journal of the American statistical Association*, *In Hemis (1999)*.
- Ula, T.A. and Smadi, A. A. (1997). Periodic stationary conditions for periodic autoregressive moving average processes as eigenvalue problems. *Water Resources. Research*. **33**, pp 1929-1934.
- Vecchia, A V. (1983). Aggregation and estimation for periodic Autoregressive-moving average models. Ph. D. Dissertation, Department of Statistics, Colorado State University, Forth Collins, Colorado.
- Vecchia, A. V., Obeysekera, J.T. and Salas, J.D. (1983). Aggregation and estimation for low-order periodic ARMA models. *Water Resources Research*, **19**, pp 1297-1306.
- Vecchia, A. V. (1985a). Maximum likelihood estimation for periodic autoregressive Moving average models. *Technometrics*, **27**, pp 375-384.

- Vecchia, A. V. (1985b). Periodic autoregressive-moving average (PARMA) modeling with application to water resources. *Water Resources Bulletin*, **21**, pp 721-730.
- Vecchia, A.V and Ballerini, R. (1992). Testing for periodic autocorrelations in seasonal time series data. *Biometrika* **78**, pp 53-63.
- Whittle, P. (1953). Estimation and information in time series, *Ark. Math. Fys. Astr.*, **2**, pp 433-434.
- Whittle, P. (1965). Recursive relations for predictors of non-stationary processes. *Journal of the Royal Statistical Society B*, **27**, pp 523-532.
- Wilson, G.T. (1979). Some efficient computational procedures for high order ARMA models, *Journal of Statistical Computational Simulation*, **8**, pp 303-309.
- Wold, H. (1938). A study in the stationary time series, Almqvist and Wiksell, Stockholm (2nd edition :1954).
- Young, P. (1984). Recursive Estimation and Time-Series Analysis, *An Introduction*. Springer-Verlag.
- Zadrozny, P. A. (1989). Analytical derivatives for estimation of linear dynamic models, *Computers and Mathematics with applications* **18**, pp 539-553.
- Zadrozny, P. A. (1992). Errata to 'Analytical derivatives for estimation of linear dynamic models, *Computers and Mathematics with applications* **24**, pp 289-290.
- Zahaf, T. (1998). "Contribution à l'estimation des paramètres de séries chronologiques". *Thèse de Doctorat es Sciences*. Institut de Statistique et de Recherche Opérationnelle, Université Libre de Bruxelles.