N° d'ordre : 04/2006-D/MT

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE D'ENSEIGNEMENT SUPERIEUR ET DE LA RECHECHE SCIENTIFIQUE
UNIVERSITE DES SCIENCES ET DE LA TECHNOLOGIE «HOUARI BOUMEDIENNE»
FACULTE DE MATHEMATIQUES

Thèse

Présentée pour l'obtention du diplôme de Doctorat

En : Mathématiques
Spécialité : Recherche Opérationnelle (Méthodes Stochastiques)
Par
AKNOUCHE Abdelhakim

Sujet

Etude théorique et algorithmique de la modélisation des séries chronologiques périodiques

Soutenue le 28/06/2006, devant le jury composé de :

Méziane	AIDER	Professeur,	U.S.T.H.B.,	Alger	Président
Mohamed	BENTARZI	Professeur,	U.S.T.H.B.,	Alger	Directeur de thèse
Amar	AISSANI	Professeur,	U.S.T.H.B.,	Alger	Examinateur
Hocine	FELLAG	Professeur,	U.M.M.T.O.,	Tizi-ouzou	Examinateur
Christian	FRANCQ	Professeur,	U. C. D., Lille3,	Lille	Examinateur
Amrane	HOUACINE	Professeur,	U.S.T.H.B.,	Alger	Examinateur
Abdelouahab	BIBI	M. C.,	U. T. M.,	Constantine	Examinateur

Theoretical and algorithmic study of periodic time series modelling

Abdelhakim Aknouche
Département de recherche opérationnelle
Faculté de Mathématiques,
Université des Sciences et de la Technologie Houari Boumediene,
U. S. T. H. B.

Remerciements

Je ne remercierais jamais assez mon Directeur de thèse et mon enseignant depuis la graduation, le Professeur Mohamed Bentarzi pour m'avoir proposé ce sujet dont les prémisses découlent de mon sujet de magister qu'il m'a proposé, il y a déjà une dizaine d'années. Cette thèse lui doit beaucoup. Sa compétence technique et son pragmatisme tout d'abord, puis sa conception de la recherche et sa sévérité scientifique, mais aussi sa passion et son abnégation pour la science m'ont conduit à tirer le meilleur de moi-même, au sein du groupe dynamisant qu'il a su créer. Qu'il trouve ici ma plus profonde admiration.

Je tiens à remercier vivement mon ancien enseignant le Professeur Meziane Aider pour m'avoir honoré de son aimable acceptation à présider le jury de cette thèse. Je trouve charmante la coïncidence qu'il y a une décennie, il a été également président de jury de mon mémoire d'ingéniorat.

Je saisis cette opportunité pour exprimer ma grande reconnaissance et ma profonde gratitude aux membres de jury, Messieurs : Amar Aissani Professeur à l'U.S.T.H.B., (qui en passant, a été président de jury de ma thèse de magister) Abdelouahab Bibi, Maître de conférences à l'université de Constantine, Hocine Fellague, Professeur à l'université de Tizi-ouzou, Christian Francq, Professeur à l'université de Lille 3, et Amrane Houacine, Professeur à l'U.S.T.H.B. pour m'avoir fait l'honneur d'accepter d'examiner cette thèse.

J'ai eu la chance de travailler avec madame Hafida Guerbyenne, mon ancienne enseignante, collègue et co-auteur avec qui j'ai beaucoup appris et dont certaines parties de ma thèse sont tirées de travaux écrits conjointement. Je voudrais lui témoigner ma profonde gratitude et mes plus vifs remerciements pour son aide précieuse, ses conseils profitables et sa contribution matérielle dans ce travail.

Je renouvelle mes remerciements et ma gratitude à Monsieur Abdelouahab Bibi, non pas en tant que membre de Jury seulement mais en ami pour ses aides précieuses, en particulier la documentation utile qu'il m'a longuement abreuvée.

Certaines parties de cette thèse sont tirées de travaux conjointement écrits avec mon collègue Fayçal Hamdi. Je n'oublierais jamais son aide et son soutien, notamment dans la programmation ainsi que pour les fastidieuses corrections.

Cette thèse n'aurait jamais été finalissée sans l'aide de monsieur Noredine Kaciabdellah, ma sœur Madame Nadjet Kaciabdellah et Monsieur Mohamed Adnane Holden qui ont tant fait pour apporter des corrections importantes sur la forme de cette thèse et des nombreux articles que j'aie soumis.

Mes collègues, chercheurs et anciens enseignants à l'U.S.T.H.B. ainsi qu'ailleurs m'ont fait apprécier l'enseignement et la recherche tout en m'encourageant et me conseillant pertinemment, merci à eux : Miloud Mihoubi (avec qui j'ai eu maintes fois des discussions enrichissantes), le professeur Abdelhafid Berrachedi chef de département de recherche opérationnelle, Souhail Chennouf, Ahmed Benselma, Ameur Abdelatif, Hacène Belbachir (pour ses aides et encouragements), Ahmed Boukhebbouze, Mohamed Said Maamra, Amar Rebbouh, Salim Laajouze, Rachid Ouafi, Ahcen Ait Haddaden, Mohamed Amine Chergui, Mustapha Moulai, Asma Bouchemakh, Khadija Djaballah, Mourad Boudhar, Djamal Chaabane, Sadek Bouroubi, Moncif Abbas et tous mes collègues du Département de Recherche Opérationnelle.

Je tiens à remercier toutes les personnes qui m'ont aidé de loin ou de près à élaborer ce travail. Je citerai en particulier : Zahir Hammoutene pour la documentation qu'il m'a longuement fournie, Aboubakr Aknouche, Abdenour Bouhouche, Mésiane Hamel, Sofiane Hamdache, Mourad Idir, Nadia Arbaoui et Lila Mentseur.

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

A ma mère
A la mémoire de mon père
A mon frère et mes deux sœurs

Alà Barakati Ellah

Contents

Introduction

Periodicity is an active area of current time series research and periodic time series models have been extensively used in the recent decades to describe the many series with periodic dynamics, encountered in a wide variety of scientific fields including environment (Jones and Brelsford, 1967; Salas et al., 1983; Vecchia 1985a, 1985b; Jymenez et al., 1989), economics (Cleveland and Tiao, 1979; Osborn, 1992; Franses, 1996; Franses and Paap, 2004) engineering (Gardner and Franks, 1975; Meyer and Burrus, 1975; Bittanti and De Nicolao, 1993; Adams and Goodwin 1995) and various other physical and natural sciences. Three basic models have been suggested in the literature, namely impulse response models with periodic structure, periodic state-space characterizations, and differential or difference equations with periodic coefficients, such as *periodic autoregressive moving average* ($PARMA$) models. The success of $PARMA$ models is due to their efficient representation of many stochastic phenomena exhibiting a periodic autocorrelation structure that cannot be adequately explained by the traditional *seasonal* $ARIMA$ ($SARIMA$) models. Another important reason for the growing interest is that they can be exploited in the analysis of *vector* $ARMA$ models ($VARMA$) in order to appreciably reduce the number of model parameters (e.g., Pagano, 1978; Newton, 1982). Moreover, such models arise usually by the discretization of linear continuous-time periodic models. However, compared to traditional seasonal characterizations, $PARMA$ models involve more complex representations including a larger number of parameters and a complex covariance structure. This makes the model building problem (identification, estimation, and diagnostic checking) as well as the study of the main properties of such models more difficult than those for traditional time-invariant seasonal models.

The inability of $SARIMA$ models suggested by Box and Jenkins (1976) to adequately represent many seasonal time series, especially those exhibiting a periodic autocovariance structure (e.g. Tiao and Grupe, 1980), has motivated the developments of research related to the class of $PARMA$ models and *periodically correlated processes*. There has been a constantly increasing interest in the building problem (identification, estimation, and diagnostic checking) of $PARMA$ models as well as in the study of some algebraic and probabilistic properties of such models, such as causality (Tiao and Grupe, 1980; Boshnakov, 1997; Hurd et al., 2002) invertibility (Čipra, 1985; Bentarzi and Hallin, 1994) spectral factorization (Bittanti and De Nicolao, 1993; Bentarzi and Hallin, 1998; Bentarzi, 1998), strict periodic stationarity, periodic ergodicity (cycloergodicity, e.g. Boyles and Gardner 1983) and the existence of higher order moments.

Since the end of the seventies, $PARMA$ modelling (identification, estimation and diagnostic checking) has seen significant interest in a wide variety of systems. The great majority of the works concerning $PARMA$ models represent extensions of similar methods for stationary $ARMA$. Several authors have studied the identification problem, in particular we mention the works of Cleveland and Tiao (1979), Sakai (1982), Vecchia (1985b), Mcleod, (1993) and Ula and Smadi (2003). These works are aimed at generalizing standard methods based on information theory, such as the AIC and BIC criteria (Sakai, 1982; Vecchia, 1985b) or those based on simple and partial autocorrelation functions (Ula and Smadi, 2003). On the other hand, the diagnostic checking stage for building periodic linear models has received much less interest. The attention has been concentrated mainly on the case of periodic autoregressions (see e.g. Vecchia and Ballerini, 1991 for the test of periodicity for $PARMA$ models and Mcleod, 1993 and 1994 for the diagnostic checking of periodic autoregressions). By studying the property of locally asymptotic normality (LAN) as well as the asymptotic linearity property, Bentarzi and Hallin (1996) have built a locally asymptotically (parametric and nonparametric) test for testing traditional autoregressions (null assumption) against their periodic counterparts.

Of particular importance in building a time series model is the parameter estimation stage. Estimation methods can be classified into two general categories according to the type of data on which they are based : off-line methods and on-line methods. In the first case, time series data are of fixed size, whereas in the second case, they are indefinitely available in real time. For $PARMA$ models, existing off-line estimation techniques have been confined to a variety of methods and approaches, such as the method of moment (Pagano, 1978; Boshnakov; 1996; Anderson and Meerschaert, 2005), the Bayesian approach (Andel, 1983), the least squares technique (Adams and Goodwin, 1995; Basawa and Lund, 2001; Bentarzi and Aknouche, 2006) and the maximum likelihood method (Vecchia, 1985a; Li and Hui, 1988; Jymenez et al., 1989; Lund and Basawa, 2000; Basawa and Lund, 2001). These techniques can be classified into two general classes, namely methods that use the data directly, and methods that apply some preliminary transformations to the data (see Porat and Friedlander 1986a for the $ARMA$ case). Among the methods in the first class, we mention in particular the exact maximum likelihood and its many approximations (conditional and unconditional) and the least squares method with its particular variants (ordinary, weighted, conditional and unconditional). Typically, these methods are aimed at preserving some desirable asymptotic properties, namely consistency, asymptotic efficiency and asymptotic normality, while reducing their computational complexity. Among the second class of methods, the most common approach, known as the method of moments, is to transform the data into a finite set of statistics and then estimate $PARMA$ parameters from these statistics. As a special case, we mention the periodic Yule-Walker method (Pagano 1978), their order-recursive Durbin-Levinson formulations (Sakai, 1982; Boshnakov, 1996), and the innovations algorithm (Anderson et al., 1999; Anderson and Meerschaert, 2005).

The off-line estimation stage for building linear periodic time series models has a long history. It has seen significant developments since the early 1960's. The literature is superabundant and the interest is increasingly growing. John and Brelsford (1967) have considered the estimation problem for particular periodic autoregressions, but without studying the asymptotic properties. After some empirical works (Bhuiya, 1971; Tao and Delleur, 1976) did not lead to consistent results, Pagano (1978) obtained interesting formal results for periodic autoregressions. The author generalized the Yule-Walker equations to the periodic case, determined the moment estimators and established their large statistical properties. After the seminal work of Pagano (1978), Cleveland and Tiao (1979) defined $PARMA$ models and showed the usefulness of such models for modelling seasonality. A similar periodically time-varying formulation, but in another context, had been proposed earlier by Meyer and Burrus (1975). When Tiao and Grupe (1980) showed the inefficiency of the $SARIMA$ models of Box and Jenkins (1970) for representing certain seasonal series with periodic autocorrelation structure, $PARMA$ modelling has seen its apogee. A considerable amount of research has been executed in the area of $PARMA$ estimation. This research includes contributions by authors such as Salas et al. (1982), Vecchia et al. (1983), Vecchia (1983) and Andel (1983). Elsewhere, Sakai (1982) improved from the algorithmic viewpoint the Yule-Walker type method of Pagano (1978) by proposing a recursive approach (with respect to the model order) based on a generalization of the Levinson-Durbin algorithm to the periodic autoregression case. Moreover, this author generalized the concept of partial autocorrelation, which represents a useful instrument for the identification stage. Hitherto, most of the works and investigations on periodic parameter estimation had concentrated mainly on periodic autoregressions. Čipra (1985), however, considered the estimation problem of periodic moving average models. This author adapted the methods of Durbin (1959) consisting of transforming a moving average model to a higher order AR model. Independently of the former works, Vecchia (1985a) proposed a more efficient algorithm based on the maximum likelihood criterion. The suggested method gives exact estimators in the case of PMA (periodic MA) models and approximate ones in the presence of a PAR (periodic AR) component. With the same approach, Li and Hui (1988) proposed an exact ML algorithm for $PARMA$ models which is statistically more efficient than that of Vecchia (1985a), but computationally cumbersome since it requires a Cholesky decomposition of matrices with a large dimension. Using also the ML criterion, Jymenez et al. (1989) developed a more efficient algorithm based on the Kalman filter. This algorithm requires, however, a considerable number of preliminary operations to start the filtering recursion. A few years later, Boshnakov (1996) combined the recursive method of Sakai (1982) for PAR models and that of Franke (1985) for $ARMA$ models to develop an order-recursive Yule-Walker algorithm for $PARMA$ parameter estimation. On the other hand, Anderson et al. (1999) generalized the innovation algorithm (cf. Kailath, 1968) for the moment estimation to the case of $PARMA$ models. The large sample properties of the periodic innovations algorithm can also be found in Anderson and Meerschaert (2005). More recently, Lund and Basawa (2000) combined the Ansley (1979) method for likelihood evaluation (for $ARMA$ models) and the periodic innovations algorithm (Anderson et al., 1999) to derive an efficient likelihood evaluation algorithm for $PARMA$ models. The proposed algorithm generalizes the one proposed by Brockwell and Davis (1988) for stationary $ARMA$ models. A possible limitation of this method is the great number of operations needed to compute the likelihood in the multivariate case. The large statistical properties of the least squares and the ML methods have been studied by Basawa and Lund (2001).

As it is easily seen, the quasi-totality of existing procedures are of an off-line nature which deal with a beforehand fixed size time series. The need to deal with a progressive size is becoming, nowadays, increasingly necessary. Indeed, in many time series application fields— such as transmission, telecommunication, signal treatment, meteorology, financial bourses and many others — the need to deal with variable series sizes is clear. In short, off-line estimation methods are inefficient when used in the presence of on-line data, mainly because their applications require, for each new available observation, start-from-scratch computations. Thus, off-line estimation algorithms are very costly in terms of the spatial and temporal complexities. For all these reasons, on-line estimation methods are appealing (or the recursive identification methods, as they are known in the engineering literature, e.g. Kushner and Clark, 1978; Ljung and Söderström, 1983; Goodwin and Sin, 1984; Young, 1984; Benveniste et al., 1990; Solo and Kong 1995; Kushner and Yin, 1997; Ljung, 1999; Belister, 2000; Pollock, 2003; Moulines et al., 2005 and further references therein), to reduce the computation time and save memory space. The main usefulness of the recursive form of on-line algorithms is due to the possibility of processing data in real time, because of a finite memorizing in a vector of fixed size. Indeed, the current estimate is updated at each new available observation so that the number of operations and the memory space do not increase with respect to the sample size, and depend only on the selected model.

Existing recursive estimation algorithms in the literature are usually associated with a particular model parametrization and the approach by which the method is built. Several approaches can be taken to derive a recursive algorithm. We mention in particular the *modified off-line*, the *Bayesian*, the *stochastic approximation*, and the *pseudo linear regression* approaches. These four approaches cover the derivation of most suggested methods (see e.g. Ljung and Söderström (1983) which contains extensive examples on this subject). However, the resulting algorithms exhibit very similar features and can be interpreted in different ways. Another way of expressing this is to say that a given recursive algorithm can be obtained from different approaches.

Despite their remarkable importance, on-line estimation methods for $PARMA$ models remain relatively unexplored, especially when compared to their stationary $ARMA$ counterparts. Indeed, it appears that the work of Adams and Goodwin (1995) is the first one to focus on this interesting problem. Adams and Goodwin (1995) adapted recursive pseudo-linear regression methods to estimate the parameters of $PARMA$ models. Moreover, they studied, based on the works of Solo (1979) and Goodwin and Sin (1984), the convergence problem of their algorithms. Unfortunately, the obtained estimators are not efficient since they are based on the ordinary least squares estimation methods. Moreover, no information about their asymptotic distribution is given.

It is evident that many physical and natural processes are more accurately modelled by a nonlinear model than by a linear representation. At present, the conditional heteroskedasticity approach for modelling stationary time series with nonlinear dynamics offers powerful analysis and design tools based on solid theoretical grounds. Such models, known as $ARCH$-type (autoregressive conditionally heteroskedastic models), have dominated the nonlinear modelling field since the appearance in 1982 of the fundamental paper of Engle (1982). Originally intended for modelling financial time series (exhibiting an instantaneous variability feature), $ARCH$-type models find by now applications of interest in many other fields such as climate (e.g. Tol, 1996), environment (see

Bordignon and Lisi, 2000 and references therein) and many others. In the last few years there has been an increased cross-fertilization (cf, Tsay, 1987; Wong and Li, 1997) of ideas between $ARCH$ -type models and other important classes of nonlinearity such as bilinear models (e.g. Tong, 1990) and random coefficient autoregressive (RCA) models (Nicholls and Quinn, 1982; Tsay, 1987; Bera et al., 1996). This has also played a crucial role in the development of the three formulations. On the other hand, several extensions of $ARCH$ models have been made in order to absorb other time series characteristics, including in particular moving average dynamics (Generalized $ARCH$: $GARCH$ models, Bollerslev, 1986), long memory feature (Ding and Granger, 1993; Baillie et al., 1996) and simultaneous dependences (multivariate $ARCH$, Ling and McAleer, 2003), while keeping stationarity. As is well known, many difficulties are encountered in developing consistent extensions to the nonstationary case (cf, Dahlhauss and Subba Rao, 2005; 2006). An important class of nonstationarity is the one associated with periodic phenomena. In fact, periodic processes arise in economics and finance often spontaneously (Bollerslev and Ghysels, 1996; Franses and Paap, 2000) or often intentionally (e.g. Bibi and Francq, 2003).

Furthermore, in the last decade, the notion of periodically correlated processes was successfully exploited in a variety of new classes of time series models, such as $ARCH$-type models (Bollerslev and Ghysels, 1996; Franses and Paap, 2000), Autoregressive Fractionally Integrated Moving Average ($ARFIMA$) models (cf., Hui and Li, 1995; Franses and Paap, 2000), periodic Markov switching models (Ghysels, 1994; Ghysels, 1997), Threshold AR models (Lewis and Bonnie, 2002) and Mixture autoregressive models (Shao, 2005). Indeed, it was observed that for many economic time series, particularly in the financial domain and monetary market (exchange rate, inflation rate, prices of economic active in a bourse, ...), many observed time series exhibit hidden periodicity in the autocovariance function, which needs the use of periodic models (cf., Abraham and Ikenberry, 1994; Bessembinder and Hertzel, 1994; Franses and Paap, 2000, Franses and Paap, 2004). Thus, the economic time series analysts have become, nowadays, more convinced for the need to include the periodicity and conditional heteroskedasticity in a time series model. In fact, the class of *Periodically correlated Generalized Autoregressive Conditionally Heteroskedastic* ($PGARCH$) models has recently received considerable interest from many economic time series analysts (see Bollerslev and Ghysels, 1996; Franses and Paap, 2000 and references therein). The motivation behind the proposition of such models for modelling economic processes, is practically justified on the one hand, by their use of the stochastic conditional variance to capture the instantaneous volatility features and, on the other hand, by the fact that they allow the underlying coefficients to be periodic time-varying to capture the periodicity in the covariance structure of the generator process. Nowadays, many financial time series display a similar dynamics to that of the data met in the adaptive control and signal processing fields. Indeed, these data (stock exchange courses, rate of exchange) are by now gradually available in small time intervals and it was theoretically proved that the $GARCH$ models are good approximations for certain continuous time series models (cf. Drost and Werker 1996; Nelson and Foster, 1994; Gourieroux et al., 1993; Nelson, 1990). Thus, the real-time estimation procedure adapted to an increasing sample size is necessary by now in such fields.

The present thesis is concerned with the study of modelling time series with periodic dynamics. We consider here two formulations, one of which is linear being expressed through stochastic difference equations with periodic (scalar and matrix) coefficients, and the second one is nonlinear of the conditionally heteroskedastic type. Our contribution is threefold :

i) First, the study of certain algebraic and probabilistic properties of such models, namely : stability (causality and invertibility), strict periodic stationarity, the autocorrelation structure, the Fisher information and the existence of higher order moments.

ii) Then, the development of estimation methods based on the maximum likelihood criterion and the least squares method. Certain methods are restricted to the univariate case and others to the multivariate one. Moreover, the estimation problem is developed according to two approaches : on-line estimation and off-line estimation.

iii) And finally, the study of the statistical properties (asymptotic) of the developed methods, including the strong consistency, the asymptotic normality and asymptotic efficiency of the estimates provided by these methods.

0.1 Presentation and contributions of the thesis

Our thesis entitled "Theoretical and algorithmic study of periodic time series modelling" comprises two parts : linear periodic modelling and nonlinear periodic modelling.

Part I : **Linear modelling of periodic time series**

Chapter 1

Chapter 1 studies some algebraic and probabilistic properties of periodic $VARMA$ ($PVARMA$)

models, such as causality, invertibility, the autocovariance structure and the $PARMA$ Fisher information matrix. We first propose some necessary and sufficient conditions for causality and invertibility, which seem computationally simpler to check than existing conditions in the literature. Moreover, we develop efficient calculation procedures for the $PVARMA$ autocovariances, based on a state space representation of such models. The proposed methods compute the autocovariances for distinct seasons separately, thereby facilitating efficient computation for models with a large period. Furthermore, the autocovariance calculation procedure will be exploited in the computation of the limiting Fisher information matrix for $PARMA$ models, through a conditional likelihood expression. Our established algorithm extends Klein and Mélard's algorithm (1989) developed for stationary $ARMA$ models. The singularity problem of such an information matrix will also be studied. Indeed, we propose a necessary and sufficient condition for nonsingular Fisher information matrix for $PARMA$ models.

Chapter 2

In this chapter we are mainly concerned with off-line estimation methods for periodic $VARMA$ models. Our study will be restricted to the class of methods that use the data directly, such as the maximum likelihood and the least squares methods. We first deal with the likelihood evaluation of $PVARMA$ models. We exploit the method for $PVARMA$ autocovariances calculation given in Chapter 1 to evaluate the $PVARMA$ likelihood by means of the Kalman filter. An alternative technique based on the Ljung and Box (1979) method for the particular class of $PARMA$ models introduced in Bentarzi (1995) is also proposed. Then, we study the large sample properties of parameter estimates for $PVARMA$ models. Our results generalize the ones obtained by Basawa and Lund (2001) to the multivariate $PARMA$ case.

Chapter 3

Chapter 3 provides an adaptation of some on-line recursive algorithms, well-known in the literature of time-invariant $ARMA$ models, to deal with $PARMA$ models. Though transforming $PARMA$ estimation problems to the ones for multivariate $VARMA$ models where several efficient on-line methods are available, the large number of parameters in the corresponding $VARMA$ models suggests direct development of a specific $PARMA$ theory would be fruitful. After reviewing some basic principles and criteria of recursive on-line estimation, we concentrate for the first time on the on-line estimation problem for PAR models. Furthermore, we provide an adaptation of the well-known recursive least squares (RLS) algorithm to deal with PAR models. We show under mild conditions that the obtained estimators are asymptotically efficient. Then, we study the general $PARMA$ recursive estimation. In particular, we generalize the obtained $PRLS$ algorithm to the $PARMA$ case. The generalization is carried out by means of two approaches, namely the pseudo-linear regression (PLR) and the recursive prediction error (RPE) method. Several variants of the PLR method (such as the extended least squares, ELS) are developed to take into account the stochastic convergence problem. The developed $PELS$ (Periodic ELS) algorithms provide strongly consistent estimators under weaker conditions than the ones given by Adams and Goodwin (1995). Moreover, we explore an alternative approach based on the recursive predictor error method ($RPEM$). In particular, we extend the well-known recursive maximum likelihood (RML) algorithm to the $PARMA$ case. The obtained estimates are shown to be asymptotically efficient. Then, we show that both $PRML$ (Periodic RML) and $PELS$ algorithms can be formulated via a single set of recursions through a certain prefilter. Moreover, it will be shown that with a modified version of this prefilter, the resulting algorithm lies somewhere in between the $PELS$ and the $PRML$. The last part of the chapter develops an improved variant of the $PRML$ algorithm, based on the Fisher information matrix. The resulting algorithm is an on-line analogue of the well-known Fisher scoring algorithm, which is devoted exclusively to the off-line case. Finally, the finite properties of the obtained estimates are studied in the last section by means of a simulation study. Furthermore, we show, through various applications of this algorithm on many $PARMA$ models, that various desirable properties take place in our periodic case.

Part II : **Nonlinear modelling of periodic time series**

Chapter 4

In this chapter, some probabilistic and algebraic properties of some periodic conditionally heteroskedastic models are investigated. The first part is concerned with the study of periodic $GARCH$ models. In particular, we review the basic definitions, concepts and assumptions concerning this class. Moreover, the causality and invertibility properties of such models are studied. Then, we explore the existence of strict periodic stationary solutions of a $PGARCH$ difference equation. Indeed, we derive a periodic Markovian state space representation from which we obtain a necessary and sufficient periodic stationarity condition for a $PGARCH$ model. Then, we obtain, under general and tractable assumptions, a necessary and sufficient condition for a general $PGARCH_S\,(q,p)$ model to have finite higher-order moments. Our established condition extends the sufficient moment condition for $PGARCH$ models obtained by Bibi (2004) and coincides, for the particular case $S=1$ (non periodic) with the necessary and sufficient condition provided by Ling and McAleer (2002) for the classical $GARCH(p,q)$. In addition, an equivalent simple necessary and sufficient moment condition for the particular and useful first order $PGARCH(1,1)$ model is given. The second part of the chapter studies the class of periodic autoregressions with period $ARCH$ errors ($PAR-PARCH$). We propose a random coefficient formulation of the $PARCH$ part in order to describe parsimoniously the varying conditional variance feature. We call the new model *double periodic AR*. The existence of a strictly periodic stationary solution and the existence of higher-order moments for such a model are studied.

Chapter 5

The main goal of this chapter is to develop some recursive estimation methods based on a two-stage least squares scheme ($2S-LS$) for estimating $PGARCH$ models. These methods, inspired from the on-line estimation litterature, (cf, Chapter 3) provide consistent and asymptotically Gaussian estimators. Firstly, we generalize the *off-line two-stage least squares* ($2S-LS$) method proposed by Bose and Mukherjee (2003) for $ARCH$ models (see also Pantula, 1988 for the $ARCH(1)$ case) to their periodic counterparts. The obtained off-line $2S-LS$ method will be adapted to a recursive on-line context. We call the resulting algorithm the *two-stage RLS* algorithm ($2S-RLS$). These two methods, besides their robustness for the conditional normality assumption, are shown to provide asymptotically Gaussian estimators. Moreover, the second method is particularly well adapted to capture the dynamic features of the financial time series that are more and more often observed on a high-frequency basis. Then, recursive estimation for $PGARCH$ models is considered. We propose two algorithms that parallel the ones available for standard linear models (cf, Chapter 3). The first one is the Two-Stage Pseudo-Linear Regression ($TS-PLR$) which is a version of the well-known Pseudo-Linear Regression (PLR) approach given in two stages; while the second one, namely the Two-Stage Recursive Predictor Error algorithm ($TS-RPE$) is an adaptation of the traditional Recursive Maximum Likelihood (RLM) method to deal with $PGARCH$ models. It will be shown that the obtained estimators are strongly consistent and in particular the RPE algorithm provides asymptotically Gaussian estimates.

Part I

Linear modelling of periodic time series

Chapter 1

Periodic $ARMA$ models and some of their properties

1.1 Introduction

Periodic linear time series models have been extensively used in the three recent decades in a variety of theoretical and practical research fields, such as hydrology (Salas et al., 1983; Vecchia 1985*a*, 1985*b*; Jymenez et al., 1989), economics (Cleveland and Tiao, 1979; Osborn, 1992; Franses, 1996; Franses and Paap, 2004), and electrical engineering (Gardner and Franks, 1975; Bittanti and De Nicolao, 1993; Adams and Goodwin 1995). The success of the latter models is due to their efficient representation of many stochastic phenomena exhibiting a periodic autocorrelation structure that cannot be adequately explained by the classical $SARIMA$ models. Another reason of the growing interest is that they can be exploited in the analysis of Vector Autoregressive Moving Average models ($VARMA$) in order to reduce appreciably the number of parameters to be taken into account (e.g., Pagano, 1978; Newton, 1982). However, compared to the $SARIMA$ formulation, *periodic* $ARMA$ ($PARMA$) models involve more complex representations, which use a much larger number of parameters and a more complex covariance structure. This makes the model building problem (identification, estimation, and diagnostic checking) and then the study of the main properties of such models more difficult than those for traditional $SARIMA$ models.

The inability of the $SARIMA$ models suggested by Box and Jenkins (1970) to adequately represent several seasonal time series, especially those exhibiting periodic autocovariance structure (e.g. Tiao and Grupe, 1980), has motivated the development of research related to the class of $PARMA$ models and *periodically correlated processes*. Several researchers were interested in the study of some algebraic properties of $PARMA$ models, such as the causality (Tiao and Grupe, 1980; Boshnakov, 1997; Hurd et al., 2002) the invertibility (Čipra, 1985; Bentarzi and Hallin, 1994) and the spectral factorization (Bittanti and De Nicolao, 1993; Bentarzi and Hallin, 1998; Bentarzi, 1998); as well as in some probabilistic properties such as the strict periodic stationarity (strict cyclostationarity), the periodic ergodicity (cycloergodicity) and the existence of higher order moments (e.g. Boyles and Gardner, 1983). For instance, the spectral factorization problem for $PARMA$ models, which consists of the construction of all possible adequate models corresponding to an autocovariance function of a certain periodically correlated process, has been completely solved by Bentarzi (1998). Similarly, the causality and invertibility problems have been widely considered in the literature of $PARMA$ models. However computationally simple necessary and sufficient conditions to check such properties remain needed in the problem of building periodic linear models.

Furthermore, most of the existing methods of building $PARMA$ models are conditioned on the causality and invertibility properties. The main difficulties involved by these methods for $PARMA$ models compared to those for their stationary counterparts lie in the periodic autocovariance structure, which does not have a symmetry and then the Toeplitz property. The main goal of this chapter is to study some algebraic and probabilistic properties of periodic $VARMA$ ($PVARMA$) models, such as causality, invertibility, the autocovariance structure and the $PARMA$ Fisher information matrix of models. We first propose some, necessary and sufficient conditions for causality and invertibility, which seem computationally simple to check. Moreover, we develop efficient calculation procedures for the $PVARMA$ autocovariances, which will be exploited in the calculation of the $PARMA$ Fisher information matrix.

The remainder of this chapter is organized as follows. In section 2, we present a succinct retrospective of the class of periodically correlated processes and their linear representation, periodic $VARMA$ models. Furthermore, we study the causality and invertibility problems of such models. Section 3 is devoted to the calculation of the $PVARMA$ autocovariances. We propose a fast algorithm based on a state space representation of the $PVARMA$ model. This algorithm works well both for univariate and multivariate models. Section 4 is mainly concerned with the calculation and the invertibility of the asymptotic Fisher information matrix of $PARMA$ models, which is a useful devise for some statistical inferences. We provide a computation algorithm based on the conditional likelihood expression. The established algorithm extends Klein and Mélard's algorithm (1989) developed for stationary $ARMA$ models. This algorithm is based on the theoretical autocovariances which can be obtained from Section 3. In addition, we give a necessary and sufficient condition for nonsingular Fisher information matrix of a $PARMA$ model.

1.2 Periodically correlated processes and periodic ARMA models

The purpose of this section is to present some basic results concerning periodic linear times series models, which proved its usefulness and appropriateness for modelling time series exhibiting periodical autocorrelation structure. We first present the strict periodicity property, which is needed for a formal theory. The more handled concept of periodic correlation (weak periodicity, cyclostationarity), which conserve the essential properties of the strict periodicity feature is then given. Finally some probabilistic and algebraic properties of periodic linear models will be studied.

1.2.1 Strictly periodic processes

Definition 1.2.1

An $m-$variate process $\{\mathbf{y}_t,\ t \in \mathbb{Z}\}$ is said to be *strictly periodic* with period S if for all $n \in \mathbb{N}^*$, $h \in \mathbb{Z}$ and $(t_1, t_2, ...t_n) \in \mathbb{Z}^n$, the joint distribution of the random vector $\left(\mathbf{y}'_{t_1}, \mathbf{y}'_{t_2}, ..., \mathbf{y}'_{t_n}\right)'$ is the same as that of $\left(\mathbf{y}'_{t_1+Sh}, \mathbf{y}'_{t_2+Sh}, ..., \mathbf{y}'_{t_n+Sh}\right)'$.

Remark 1.2.1

Clearly, the strict periodicity concept generalizes that of strict stationarity. Indeed, from Definition 1.2.1 it follows that a strictly periodic process with period $S = 1$ is strictly stationary.

A particular and useful subclass of strict periodic processes is the class of independent and periodically distributed ($i.p.d.$) random variables.

Definition 1.2.2

A random sequence $\{\boldsymbol{\varepsilon}_t,\ t \in \mathbb{Z}\}$ is said to be *independent and periodically distributed* random variables with period S, denoted here by ($i.p.d.$), if :

$i)$ $\{\boldsymbol{\varepsilon}_t,\ t \in \mathbb{Z}\}$ is a sequence of independent random variables,

$ii)$ the probability distribution of $\boldsymbol{\varepsilon}_t$ is the same as that of $\boldsymbol{\varepsilon}_{t+hS}$ for all $t, h \in \mathbb{Z}$.

Obviously, an ($i.p.d.$) sequence of period S is a strictly S-periodic process. Moreover, for $S = 1$ it coincides with an independent and identically distributed random ($i.i.d.$) sequence.

Furthermore, the following proposition shows that any mean square convergent combination (with S-periodic coefficients) of S-periodic (*i.p.d.*) sequence is a strictly S-periodic process.

Proposition 1.2.1 (Aknouche and Guerbyenne, 2005)

Let $\{\boldsymbol{\varepsilon}_t,\ t \in \mathbb{Z}\}$ *be (i.p.d.) random sequence with period* S *and define the random process* $\{\mathbf{y}_t,\ t \in \mathbb{Z}\}$ *by*

$$\mathbf{y}_t = \sum_{j=0}^{\infty} \psi_{tj} \boldsymbol{\varepsilon}_{t-j},$$

where $\left\{\psi_{tj}\right\}_j$ *is a sequence of absolutely summable* $S-$*periodic matrix coefficients, i.e.* $\psi_{t+Sh,j} = \psi_{tj}$ *for any* $t, h \in \mathbb{Z}$. *Then, the process* $\{\mathbf{y}_t,\ t \in \mathbb{Z}\}$ *is strictly periodic with period* S.

Proof

The proof can be straightforwardly established by showing that the characteristic function of the vector $\mathbf{Y}(h,n) = \left(\mathbf{y}'_{t_1+Sh}, \mathbf{y}'_{t_2+Sh}, ..., \mathbf{y}'_{t_n+Sh}\right)'$ is the same as that of $\mathbf{Y}(n) = \left(\mathbf{y}'_{t_1}, \mathbf{y}'_{t_2}, ..., \mathbf{y}'_{t_n}\right)'$ for any $n \in \mathbb{N}^*$, $h \in \mathbb{Z}$ and $(t_1, t_2, ...t_n) \in \mathbb{Z}^n$. This can be easily shown using the representation $\mathbf{y}_t = \sum_{j=0}^{\infty} \psi_{tj} \boldsymbol{\varepsilon}_{t-j}$, the S-periodicity of the sequence $\left\{\psi_{tj}\right\}$ and the (*i.p.d.*) property of the process $\{\boldsymbol{\varepsilon}_t,\ t \in \mathbb{Z}\}$. ■

As for strict stationarity, it is also practically impossible to check the strict periodicity in stochastic processes except in very special cases. However, many of the important statistical properties of stochastic processes can be obtained from their first and second order moment (if these exist) ; therefore, periodicity in these two moments can be enough to explain the periodic property of the process. These facts lead to the concept of periodic correlation.

1.2.2 Periodically correlated processes

The concept of periodic correlation introduced by Gladyshev (1961) was the cornerstone of the class of periodic linear autoregressive moving average models. The second order process $\{\mathbf{y}_t,\ t \in \mathbb{Z}\}$ is said to be *periodically correlated* (weakly periodic, cyclostationary) if there exists a positive integer S such that the mean and the autocovariance function are S-periodic in time, that is

$$\begin{array}{l} i)\ \mu_{t+S\tau} = \mu_t \\ ii)\ \boldsymbol{\gamma}_h^{(t+S\tau)} = \boldsymbol{\gamma}_h^{(t)} \end{array} \qquad \forall\ t, h, \tau \in \mathbb{Z},$$

where $\mu_t = E(\mathbf{y}_t)$ is the mean of the process and $\boldsymbol{\gamma}_h^{(t)} = E(\mathbf{y}_t\mathbf{y}_s')$ is the autocovariance (matrix) function at time t and lag h. By periodicity, this last one can be seen as a set of S functions. As basic properties of these functions we have

$$\boldsymbol{\gamma}_h^{(s+S\tau)} = \boldsymbol{\gamma}_h^{(s)},$$

and

$$\boldsymbol{\gamma}_{-h}^{(s)} = \left(\boldsymbol{\gamma}_h^{(s+h)}\right)'.$$

In his seminal article, Gladyshev (1961) has given a one to one correspondence between univariate S-periodically correlated processes and S-variate second order stationary processes. This relation allows the study of some basic properties of an m-variate S-periodically correlated process from its equivalent Sm-variate stationary process, where many theoretical results have been already explored.

Theorem 1.2.1 (Gladyshev, 1961)

Let $\{\mathbf{y}_t,\ t \in \mathbb{Z}\}$ be an m-variate periodically correlated process of period S from which is defined the Sm-variate process $\{X(\tau),\ \tau \in \mathbb{Z}\}$ by

$$\begin{aligned} X(\tau) &= \left(X_1'(\tau), X_2'(\tau), ..., X_S'(\tau)\right)', \\ &= \left(\mathbf{y}_{1+S\tau}', \mathbf{y}_{2+S\tau}', ..., \mathbf{y}_{S+S\tau}'\right)'. \end{aligned}$$

Then, $\{\mathbf{y}_t,\ t \in \mathbb{Z}\}$ is periodically correlated if and only if $\{X(\tau),\ \tau \in \mathbb{Z}\}$ is second order stationary.

■

A very simple and important periodically correlated process is the so-called *periodic white noise.* We recall that a periodic white noise $\{\boldsymbol{\varepsilon}_t, t \in \mathbb{Z}\}$ is a zero mean sequence of uncorrelated random variables with a periodic covariance matrix $E\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_{t-h}'\right) = \delta_{h0}\Sigma_t$ where δ denotes the usual Kronecker function. In what follows, an (*i.p.d.*) sequence will be called strongly periodic white noise.

As for stationarity, a very wide class of periodically stationary processes can be generated by using periodic white noise as a general term in a set of linear difference equations. This fact was theoretically motivated by the well-known Wold-Cramér theorem (established by Cramér (1961) which has generalized the Wold (1938) theorem to nonstationary processes), which states that any stochastic process (either stationary or not) with finite variance can be decomposed as a mean square convergent sum of the past and the present of an innovation sequence. The latter Wold-Cramér result was the origin of the so-called time varying $ARMA$ models (see e.g. Hallin 1984; 1986; Dahlhauss, 1997) which are approximations of the Wold-Cramér decomposition. A particular class of time varying $ARMA$ models which is useful for representing stochastic phenomena exhibiting periodic structure is the class of periodic $ARMA$ models whose coefficients are periodic-time-varying.

1.2.3 Periodic $VARMA$ models

A second order $m-$variate periodically correlated process $\{\mathbf{y}_t,\ t \in \mathbb{Z}\}$ is said to have a periodic vector autoregressive representation of orders p_t and q_t and period S, denoted by $PVARMA_S\left(p_t, q_t\right)$, if it is a solution of a linear stochastic difference equation of the form

$$\mathbf{y}_t - \sum_{j=1}^{p_t} \boldsymbol{\phi}_{t,j}\mathbf{y}_{t-j} = \boldsymbol{\varepsilon}_t - \sum_{j=1}^{q_t} \boldsymbol{\theta}_{t,j}\boldsymbol{\varepsilon}_{t-j}, \tag{1.2.1}$$

where $\{\boldsymbol{\varepsilon}_t, t \in \mathbb{Z}\}$ is an $m-$variate periodic white noise process with covariance matrix $E\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right) = \Sigma_t$. The $m \times m$ matrix coefficients $\boldsymbol{\phi}_{t,j},\ j = 1, ..., p$ and $\boldsymbol{\theta}_{t,i},\ i = 1, ..., q,$ the white noise covariance matrix Σ_t and the orders p_t and q_t are $S-$periodic in time. The $PVARMA$ model (1.2.1) is usually written in the symbolic form

$$\boldsymbol{\phi}_t(L)\mathbf{y}_t = \boldsymbol{\theta}_t(L)\boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{\phi}_t(L) = \mathbf{I}_m - \sum\limits_{j=1}^{p_t} \boldsymbol{\phi}_{t,j}L^j$, $\boldsymbol{\theta}_t(L) = \mathbf{I}_m - \sum\limits_{j=1}^{q_t} \boldsymbol{\theta}_{t,j}L^j$ and L is the backshift operator.

Note that the latter $PVARMA_S\left(p_t, q_t\right)$ model contains the periodic autoregression $PVAR_S\left(p_t\right)$ model

$$\mathbf{y}_t - \sum_{j=1}^{p_t} \boldsymbol{\phi}_{t,j}\mathbf{y}_{t-j} = \boldsymbol{\varepsilon}_t, \tag{1.2.2}$$

and the periodic moving average $P - VMA_S\left(q_t\right)$ model

$$\mathbf{y}_t = \boldsymbol{\varepsilon}_t - \sum_{j=1}^{q_t} \boldsymbol{\theta}_{t,j}\boldsymbol{\varepsilon}_{t-j}, \tag{1.2.3}$$

as special cases when $q_t = 0$ and $p_t = 0$ respectively.

For convenience, for the particular case $m=1$, the univariate $PVARMA$ process $\{y_t,\ t\in\mathbb{Z}\}$, denoted by $PARMA$, will be written in this work as follows

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

where $\{\varepsilon_t,t\in\mathbb{Z}\}$ is a univariate periodic white noise process with variance $E\left(\varepsilon_t^2\right)=\sigma_t^2$.

Based on the aforementioned Gladyshev theorem (1961), which ensures that any S-periodically correlated process can be represented equivalently as a stationary $S-$dimension process, one can rewrite an m-variate S-periodic $PVARMA$ model with order (p_t,q_t) in an equivalent stationary mS-variate $VARMA$ model with orders (p^*,q^*) obtained from (p_t,q_t). Letting $t=i+\tau S$, where $\tau=\left\lceil\frac{t-1}{S}\right\rceil$, $i=1,...,S$ and t, $\tau\in\mathbb{Z}$, the periodic model (1.2.1) can then be written in the following form, which will be used very often in the subsequent sections

$$\mathbf{y}_{i+\tau d}-\sum_{j=1}^{p_i}\boldsymbol{\phi}_{i,j}\,\mathbf{y}_{i-j+\tau d}=\boldsymbol{\varepsilon}_{i+\tau d}-\sum_{j=1}^{q_i}\boldsymbol{\theta}_{i,j}\,\boldsymbol{\varepsilon}_{i-j+\tau d}.$$

For the parsimony problem, we suppose that the polynomial operators $\boldsymbol{\phi}_t\left(L\right)$ and $\boldsymbol{\theta}_t\left(L\right)$ have no common roots. Thus, the former equation can be written in the Sm-variate $VARMA$ form (e.g. Tiao and Grupe 1980; Vecchia 1985; Bentarzi, 1995)

$$\underline{\Phi}_0\underline{X}_\tau-\sum_{k=1}^{p^*}\underline{\Phi}_k\underline{Y}_{\tau-k}=\underline{\Theta}_0\underline{\epsilon}_\tau-\sum_{k=1}^{q^*}\underline{\Theta}_k\underline{\epsilon}_{\tau-k}, \tag{1.2.4}$$

where $\underline{Y}_i=\left(\mathbf{y}_{1+(i-1)S},\mathbf{y}_{2+(i-1)S},...\mathbf{y}_{S+(i-1)S}\right)'$, $\underline{\epsilon}_i=\left(\boldsymbol{\varepsilon}_{1+(i-1)S},\boldsymbol{\varepsilon}_{2+(i-1)S},...,\ \boldsymbol{\varepsilon}_{S+(i-1)S}\right)'$, $p^*=\max_{1\leq i\leq S}\left\lceil\left(p_i-i\right)/S\right\rceil+1$, $q^*=\max_{1\leq i\leq S}\left\lceil\left(q_i-i\right)/S\right\rceil+1$, where $\lceil x\rceil$ denotes the smallest integer greater than or equal to x. The matrices of autoregressive and moving average coefficients are given by

$$(\underline{\Phi}_0)_{ij}=\begin{cases}\mathbf{I}_m & i=j\\ \mathbf{0}_m & i<j\\ \boldsymbol{\phi}_{i,i-j} & i>j\end{cases};\ (\underline{\Phi}_k)_{ij}=\boldsymbol{\phi}_{i,kS+i-j}\ ,\quad 1\leq k\leq p^*\ ,$$

and

$$(\underline{\Theta}_0)_{ij}=\begin{cases}\mathbf{I}_m & i=j\\ \mathbf{0}_m & i<j\\ \boldsymbol{\theta}_{i,i-j} & i>j\end{cases};\ (\underline{\Theta}_k)_{ij}=\boldsymbol{\theta}_{i,kS+i-j}\ ,\quad 1\leq k\leq q^*.$$

Note that the elements of the nonsingular matrices $\underline{\Phi}_0$ and $\underline{\Theta}_0$ are such that

$$(\underline{\Phi}_k)_{ij} = 0, \text{ for } k = 1, ..., p^* \text{ and } j < i - p_i + Sk,$$

and

$$(\underline{\Theta}_k)_{ij} = 0, \text{ for } k = 1, ..., q^* \text{ and } j < i - q_i + Sk.$$

1.2.4 A particular class of $PARMA$ models

This subsection is devoted to studying an important particular subclass of the family of univariate $PARMA$ models. Assuming that the orders p_t and q_t and the period S satisfy some relations which are widely encountered in practice, one can show that the corresponding univariate $PARMA$ can be written as a stationary $VARMA$ model of order $(1, 1)$. This last equivalent form enables to investigate deeply many properties and to obtain many explicit results concerning in particular the autocorrelation structure, the Wold representation and the deconvolution formula. This class of models has been introduced by Bentarzi (1995). Let $\Im_{11}$ be the class of $PARMA$ models whose corresponding S-variate $VARMA$ with equation (1.2.4) is of order $(1, 1)$. Then a $PARMA$ model is an element of $\Im_{11}$ if and only if

$$1 \leq p_0 \leq S \text{ and } 1 \leq q_0 \leq S, \tag{1.2.5}$$

where $p_0 = \max_{1 \leq i \leq S}(p_i - i + 1)$ and $q_0 = \max_{1 \leq i \leq S}(q_i - i + 1)$.

Consider a univariate $PARMA$ model from the class $\Im_{11}$

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

where its corresponding S-variate $VARMA$ is defined by the following equation (cf, equation (1.2.4))

$$\underline{\Phi}_0 \underline{X}_\tau - \underline{\Phi}_1 \underline{X}_{\tau-1} = \underline{\Theta}_0 \underline{\epsilon}_\tau - \underline{\Theta}_1 \underline{\epsilon}_{\tau-1}$$

This can also be written as

$$\underline{X}_\tau - \underline{\Phi} \underline{X}_{\tau-1} = \underline{\mathbf{\Theta}}_0 \underline{\epsilon}_\tau - \underline{\mathbf{\Theta}}_1 \underline{\epsilon}_{\tau-1} \tag{1.2.6}$$

where $\underline{\Phi} = \underline{\Phi}_0^{-1} \underline{\Phi}_1$, $\underline{\mathbf{\Theta}}_0 = \underline{\Phi}_0^{-1} \underline{\Theta}_0$, and $\underline{\mathbf{\Theta}}_1 = \underline{\Phi}_0^{-1} \underline{\Theta}_1$. Note that since the $S - p_0$ first columns of the matrix $\underline{\Phi}_1$ are null, it will be the same for the matrices $\underline{\Phi}$, $\underline{\mathbf{\Theta}}_0$ and $\underline{\mathbf{\Theta}}_1$. Hence

$$\begin{aligned}
&\underline{\Phi} = \begin{pmatrix} 0_{(S-p_0)\times(S-p_0)} & \widetilde{\underline{\Phi}}_{(S-p_0)\times p_0} \\ 0_{p_0\times(S-p_0)} & \underline{\Phi}^*_{p_0\times p_0} \end{pmatrix}, \; \underline{\mathbf{\Theta}}_0 = \begin{pmatrix} 0_{(S-p_0)\times(S-p_0)} & \widetilde{\underline{\Theta}}_{0(S-p_0)\times p_0} \\ 0_{p_0\times(S-p_0)} & \underline{\Theta}^*_{0p_0\times p_0} \end{pmatrix} \\
&\underline{\mathbf{\Theta}}_1 = \begin{pmatrix} 0_{(S-p_0)\times(S-p_0)} & \widetilde{\underline{\Theta}}_{1(S-p_0)\times p_0} \\ 0_{p_0\times(S-p_0)} & \underline{\Theta}^*_{1p_0\times p_0} \end{pmatrix},
\end{aligned} \tag{1.2.7}$$

and it is easy to verify that

$$\underline{\Phi}^m = \begin{pmatrix} 0 & \widetilde{\underline{\Phi}}(\underline{\Phi}^*)^{m-1} \\ 0 & (\underline{\Phi}^*)^m \end{pmatrix}. \tag{1.2.8}$$

Let $\underline{\Theta} = \underline{\Theta}_0^{-1}\underline{\Theta}_1$. Because the $S - q_0$ first columns of the $S \times S$-matrix $\underline{\Theta}$ are null, (q_0 being given by (1.2.5)), the matrix $\underline{\Theta}$ has then the form

$$\underline{\Theta} = \begin{pmatrix} 0_{(S-p_0)\times(S-p_0)} & \widetilde{\underline{\Theta}}_{(S-p_0)\times p_0} \\ 0_{p_0\times(S-p_0)} & \underline{\Theta}^*_{p_0\times p_0} \end{pmatrix}. \tag{1.2.9}$$

As for (1.2.8), we can deduce a similar expression for $\underline{\mathbf{\Theta}}_0^m$, $\underline{\mathbf{\Theta}}_1^m$, and $\underline{\Theta}^m$. This can be exploited to reduce the calculation of $\underline{\Phi}^m$ to that of $(\underline{\Phi}^*)^m$, which has a lower dimension. In section 2.2 we concentrate on the likelihood expression of a $PARMA_S\,(p_t, q_t)$ with an associated multivariate $VARMA\,(1,1)$.

One of the most important properties of $PVARMA$ models are the causality and invertibility. We recall that a $PVARMA$ model is causal if it can be expressed as a mean square convergent sum of the past of the innovation process. Similarly, it is called invertible if the innovation process can be given as a mean square convergent sum of the past of the underlying process. In the following subsection we study such properties.

1.2.5 Causality and invertibility of $PVARMA$ models

We first study the causality (periodic stationarity) problem of model (1.2.1).

Causality conditions

As causality involves only the autoregressive part, we consider the $PVAR$ model given by (1.2.2). A necessary and sufficient periodic stationarity condition for an S-periodic $m-$variate $PVAR$ model is frequently established using the so-called "*Period Span Lumping*" approach of Gladyshev (1961).

This approach consists of embedding the underlying S-periodically correlated m-variate process into an mS-variate stationary process and then studying the properties of this one (cf. Tiao and Grupe, 1980; Vecchia, 1985; Ula and Smadi, 1997; Lund and Basawa, 1999). Considering now a backward representation $\underline{Y}_n = \left(\mathbf{y}'_{S+nS}, ..., \mathbf{y}'_{2+nS}, \mathbf{y}'_{1+nS}\right)'$ and $\underline{\varepsilon}_n = (\varepsilon'_{S+nS}, ..., \varepsilon'_{2+nS},\ \varepsilon'_{1+nS})'$, model (1.2.2) can be rewritten in the mS-variate VAR form

$$\Phi_0 \underline{X}_\tau - \sum_{k=1}^{p^*} \Phi_k \underline{X}_{\tau-k} = \underline{\varepsilon}_\tau, \tag{1.2.10}$$

where $p^* = \max_{1\leq s\leq S} \lceil (p_s - s)/S \rceil + 1$, with $\lceil x \rceil$ denotes the smallest integer greater than or equal to x. The autoregressive coefficients are given by

$$(\Phi_0)_{ij} = \begin{cases} \mathbf{I}_m & i = j \\ \mathbf{0}_m & i > j \\ -\phi_{i,i-j} & i < j \end{cases}$$

and

$$(\Phi_k)_{ij} = \phi_{i,kS+i-j},\ i = j = 1, 2, ...S \text{ and } 1 \leq k \leq p^*,$$

where $\mathbf{0}_m$ denotes the null square matrix of order m. Therefore, the periodic stationarity condition of the model (1.2.2) is the same as the stationarity condition of its equivalent VAR representation (1.2.4b); that is, the roots of the determinantal equation

$$\det\left(\mathbf{I}_m z^{p^*} - \sum_{k=1}^{p^*} \Phi_0^{-1}\Phi_k z^{p^*-k}\right) = 0, \tag{1.2.11}$$

are less than unity in modulus (e.g., Hannan 1970, Fuller 1976). Moreover, Ula and Smadi (1997), have simplified the latter condition for model (1.2.10) by means of its corresponding state space representation. Indeed, they showed that the roots of (1.2.11) are exactly the eigenvalues of the matrix appearing in the obtained first order state space representation

$$\mathbf{\Phi} = \begin{pmatrix} \Phi_0^{-1}\Phi_1 \ldots \Phi_0^{-1}\Phi_{p^*-1} & \Phi_0^{-1}\Phi_{p^*} \\ \mathbf{I}_{m(p^*-1)\times m(p^*-1)} & \mathbf{0}_{m(p^*-1)\times m} \end{pmatrix}.$$

Consequently, model (1.2.10) is causal if and only if the eigenvalues of Φ are less than 1 in modulus.

On the other hand, using the so-called "*order span lumping*" approach introduced by Bentarzi and Hallin (1994), in which they considered a lumping of the periodic process over a span of order

p (cf. Bentarzi and Hallin, 1994; Ula and Smadi, 1997; Bentarzi, 1998), we can also obtain a necessary and sufficient causality condition for a periodic m-variate $PVAR$ model. Indeed, letting $\mathbf{Y}_{pT} = \left(\mathbf{y}'_{pT}, \mathbf{y}'_{pT-1}, ..., \mathbf{y}'_{pT-p+1}\right)'$ and $\underline{\boldsymbol{\varepsilon}}_{pT} = \left(\boldsymbol{\varepsilon}'_{pT}, \boldsymbol{\varepsilon}'_{pT-1}, ..., \boldsymbol{\varepsilon}'_{pT-p+1}\right)'$ where $p = \max_{1\leq s\leq S}\{p_s\}$, one can rewrite (1.2.10) as a first order periodic mS-variate $PVAR$ model

$$\mathbf{A}_{T,0}Y_{Tp} + \mathbf{A}_{T,1}Y_{Tp-1} = \underline{\boldsymbol{\varepsilon}}_{Tp}, \tag{1.2.12}$$

where the coefficient $\mathbf{A}_{T,0}$ and $\mathbf{A}_{T,1}$ are given by

$$(\mathbf{A}_{T,0})_{ij} = \begin{cases} \mathbf{I}_m & i = j \\ \mathbf{0}_m & i > j \\ -\phi_{Tp-i+1,j-i} & j < i \end{cases};$$

and

$$(\mathbf{A}_{T,1})_{ij} = \begin{cases} \mathbf{I}_m & i = j \\ \mathbf{0}_m & i < j \\ -\phi_{Tp-i+1,p-i+j} & i > j \end{cases},$$

are periodic with period $\mathbb{S}$, such that $p\mathbb{S}$ is the least common multiple of p and S. From representation (1.2.12), which is equivalent to (1.2.10), one can easily verify that a necessary and sufficient causality condition for model (1.2.10) is that the eigenvalues of the matrix

$$\mathbf{Q} = \mathbf{A}_{\mathbb{S},0}^{-1}\mathbf{A}_{\mathbb{S},1}...\mathbf{A}_{2,0}^{-1}\mathbf{A}_{2,1}\mathbf{A}_{1,0}^{-1}\mathbf{A}_{1,1},$$

are all less than 1 in modulus (see Ula and Smadi, 1997; Bentarzi and Hallin, 1994).

In what follows, we use an alternative approach, based on a state space representation of (1.2.10), to obtain an equivalent necessary and sufficient causality condition for model (1.2.2). This condition seems simpler than the two previous conditions.

Define the mp-variate random vectors $\mathbf{Y}_t = (\mathbf{y}'_t, \mathbf{y}'_{t-1}, ..., \mathbf{y}'_{t-p+1})'$ and $\boldsymbol{\eta}_t = (\varepsilon'_t, \mathbf{0}'_{m(p-1)\times 1})'$ and the S-periodic matrix $\mathbf{\Phi}_t$ of dimension $mp \times mp$ by

$$\mathbf{\Phi}_t = \left(\begin{array}{c|c} \phi_{t,1}\cdots\phi_{t,p-1} & \phi_{t,p} \\ \hline \mathbf{I}_{m(p-1)\times m(p-1)} & \mathbf{0}_{m(p-1)\times m} \end{array}\right).$$

Then one can rewrite model (1.2.10) in the following equivalent S-periodic state space representation, which will be used in the subsequent sections

$$\mathbf{Y}_t = \mathbf{\Phi}_t\mathbf{Y}_{t-1} + \boldsymbol{\eta}_t. \tag{1.2.13}$$

This representation was first exploited by Lütkepohl (1991, Chapter 12) for studying both the causality (stability) and the maximum likelihood estimation problems for $PVAR$ models, via the Gladyshev's (1961) period span lumping approach.

Let $\mathbf{\Omega} = \prod_{s=1}^{S} \mathbf{\Phi}_{S-s+1}$. Then the following proposition gives a necessary and sufficient causality condition for model (1.2.13) and hence for model (1.2.10).

Proposition 1.2.2 (Aknouche, 2006)

The $m-$variate, $S-$periodic $PVAR$ model (1.2.2) *is causal if and only if*

$$\rho\left(\mathbf{\Omega}\right) < 1, \tag{1.2.14}$$

where $\rho\left(A\right)$ denotes, as usual, the spectral radius of the matrix A, i.e. the maximum eigenvalue in modulus of the matrix A.

Proof

It is easy to verify that the solution of the homogeneous first order difference equation associated to the first order autoregressive operator of (1.2.13) is given by

$$G\left(t,t_0\right) = \mathbf{\Phi}_t \, \mathbf{\Phi}_{t-1} \ldots \mathbf{\Phi}_{t_0+2} \, \mathbf{\Phi}_{t_0+1}.$$

This solution corresponds to the initial value $G\left(t_0,t_0\right) = \mathbf{I}_{m\times m}$, where t_0 is the starting time of the process. The general solution of the nonhomogeneous first order stochastic difference equation (1.2.13) is then given by

$$\mathbf{Y}_t = \sum_{j=t_0+1}^{t} G\left(t,j\right) \, \boldsymbol{\eta}_j + G\left(t,t_0\right) \mathbf{Y}_{t_0}.$$

By letting t_0 decrease without limit we have

$$\lim_{t_0\to-\infty} \left\| \mathbf{Y}_t - \sum_{j=t_0+1}^{t} G\left(t,j\right) \boldsymbol{\eta}_j \right\| = \lim_{t_0\to-\infty} \left\| G\left(t,t_0\right) \mathbf{Y}_{t_0} \right\|,$$

where $\|.\|$ stands for the Euclidean norm. Thus, a necessary and sufficient causality condition for model (1.2.13) can be obtained by applying Hallin's Theorem (Hallin, 1986, Theorem 3.1) to our present particular periodic case (see also Bentarzi and Hallin, 1994); that is,

$$\lim_{t_0\to-\infty} \left\| G\left(t,t_0\right) \right\| = 0 \text{ for some fixed } t. \tag{1.2.15}$$

Let $t-t_0 = s_0 + n_0 S$, for $s_0 \in \{0, ..., S-1\}$ and $n_0 \in \mathbb{Z}$. Using the periodicity of $\mathbf{\Phi}_t$, one can write $G(t, t_0)$ in the form

$$G(t, t_0) = \left(\mathbf{\Phi}_{t_0+s_0}\ \mathbf{\Phi}_{t_0+s_0-1} ....\ \mathbf{\Phi}_{t_0+s_0-S+1}\right)^{n_0} \mathbf{\Phi}_{t_0+s_0}\ \mathbf{\Phi}_{t_0+s_0-1} .... \mathbf{\Phi}_{t_0+2}\ \mathbf{\Phi}_{t_0+1}.$$

Since $\|\mathbf{\Phi}_{s_0+t_0} \mathbf{\Phi}_{s_0+t_0-1} ... \mathbf{\Phi}_{t_0+2}\ \mathbf{\Phi}_{t_0+1}\|$ is uniformly bounded by say

$$\max_{1 \le s \le S} \|\mathbf{\Phi}_s\ \mathbf{\Phi}_{s-1} ... \mathbf{\Phi}_{s-S+1}\| = C,$$

then the left member in (1.2.15) can be rewritten in the form

$$\lim_{t_0 \to -\infty} \|G(t, t_0)\| \le C \lim_{n_0 \to \infty} \|\mathbf{\Phi}_{s_0}\ \mathbf{\Phi}_{s_0-1}\ .... \mathbf{\Phi}_{s_0-S+2}\ \mathbf{\Phi}_{s_0-S+1}\|^{n_0}$$

from which, by a cyclical rearrangement, follows the mentioned necessary and sufficient condition. ■

Similar arguments for a univariate first order PAR model can be found in Hurd et al. (2002). Thus, under (1.2.14), the $PVAR$ process $\{\mathbf{y}_t, t \in \mathbb{Z}\}$ given by (1.2.10) or equivalently by (1.2.13) admits the following two equivalent Wold-Cramér representations

$$\begin{aligned} \mathbf{y}_t &= \sum_{j=0}^{\infty} \boldsymbol{\psi}_{t,j} \boldsymbol{\varepsilon}_{t-j}, \qquad (1.2.16) \\ \mathbf{Y}_t &= \sum_{j=0}^{\infty} \mathbf{\Psi}_{t,j} \boldsymbol{\eta}_{t-j}, \end{aligned}$$

where the S-periodic $\boldsymbol{\psi}_{t,j}$ are obtained recursively via the equations

$$\left\{ \begin{array}{l} \boldsymbol{\psi}_{s,0} = \mathbf{I}_{m \times m} \\ \boldsymbol{\psi}_{s,k} = \sum\limits_{j=1}^{\min(k,p)} \boldsymbol{\phi}_{s,j} \boldsymbol{\psi}_{s-j,k-j} \end{array} \right. , \ s = 1, ..., S,\ k \in \mathbb{N},$$

and the S-periodic coefficients $(\mathbf{\Psi}_{t,j})$ are explicitly given by

$$\mathbf{\Psi}_{t,j} = \mathbf{\Phi}_t\ \mathbf{\Phi}_{t-1}\ ....\ \mathbf{\Phi}_{t-j}, \quad j \in \mathbb{N}, t \in \mathbb{Z}.$$

From the above, we see that the three studied approaches are similar in essence; that is, they reduce the determination of periodic stationarity conditions to eigenvalue problems. However, it seems that our proposed approach is much simpler than the others. To see the relations between the results of these different approaches, we consider some examples.

Examples

i) For the $PVAR_2\,(3)$ model of period 2 and order 3 given in Ula and Smadi (1997, p. 1933), and by means of which they noticed that Φ is analytically much simpler than $\boldsymbol{\Theta}$, we have

$$\boldsymbol{\Omega} = \begin{pmatrix} \phi_{2,2}+\phi_{2,1}\phi_{1,1} & \phi_{2,3}+\phi_{2,1}\phi_{1,2} & \phi_{2,1}\phi_{1,3} \\ \phi_{1,1} & \phi_{1,2} & \phi_{1,3} \\ \mathbf{I}_m & \mathbf{0}_m & \mathbf{0}_m \end{pmatrix},$$

$$\boldsymbol{\Phi} = \begin{pmatrix} & & & \mathbf{0}_m \\ & \boldsymbol{\Omega} & & \mathbf{0}_m \\ & & & \mathbf{0}_m \\ \mathbf{0}_m & \mathbf{I}_m & \mathbf{0}_m & \mathbf{0}_m \end{pmatrix},$$

$$\mathbf{Q} = \begin{pmatrix} \mathbf{I_m} & \phi_{2,1} & \phi_{2,2}+\phi_{2,1}\phi_{1,1} \\ \mathbf{0}_m & \mathbf{I}_m & \phi_{1,1} \\ \mathbf{0}_m & \mathbf{0}_m & \mathbf{I}_m \end{pmatrix} \begin{pmatrix} \phi_{2,3} & \mathbf{0}_m & \mathbf{0}_m \\ \phi_{1,2} & \phi_{1,3} & \mathbf{0}_m \\ \phi_{2,1} & \phi_{2,2} & \phi_{2,3} \end{pmatrix} \times$$
$$\begin{pmatrix} \mathbf{I}_m & \phi_{1,1} & \phi_{1,2}+\phi_{1,1}\phi_{2,1} \\ \mathbf{0}_m & \mathbf{I}_m & \phi_{2,1} \\ \mathbf{0}_m & \mathbf{0}_m & \mathbf{I}_m \end{pmatrix} \begin{pmatrix} \phi_{1,3} & \mathbf{0}_m & \mathbf{0}_m \\ \phi_{2,2} & \phi_{2,3} & \mathbf{0}_m \\ \phi_{1,1} & \phi_{1,2} & \phi_{1,3} \end{pmatrix}$$

and therefore $\boldsymbol{\Omega}$ is equivalent to $\boldsymbol{\Phi}$.

ii) For the $PVAR_2(1)$ model, we have $\boldsymbol{\Omega} = \mathbf{Q} = \phi_{2,1}\phi_{1,1}$ and $\boldsymbol{\Phi} = \begin{pmatrix} \phi_{2,1}\phi_{1,1} & \mathbf{0}_m \\ \phi_{1,1} & \mathbf{0}_m \end{pmatrix}$.

iii) Finally with a $PVAR_2\,(2)$ model of order 2 and period 2, we have $\boldsymbol{\Omega} = \boldsymbol{\Phi} = \mathbf{Q} = \begin{pmatrix} \phi_{2,1}\phi_{1,1}+\phi_{2,2} & \phi_{2,1}\phi_{1,2} \\ \phi_{1,1} & \phi_{1,2} \end{pmatrix}$. Here, the three matrices require the same level of difficulty to construct.

Remark 1.2.2

Another advantage of (1.2.14) is that in the case where $\{\varepsilon_t\}$ are independent and periodically distributed with finite higher order moment, it is also a sufficient condition for the existence of higher order moments for the $PVARMA$ process $\{\mathbf{y}_t,\ t \in \mathbb{Z}\}$.

Invertibility conditions

The invertibility problem is closely related to that of causality since it involves similar mathematical techniques. For simplicity and without loss of generality, we consider the moving average model (1.2.3). A similar condition for invertibility can be easily obtained by replacing in $\boldsymbol{\Phi}_t$ (given in

(1.2.13)) the autoregressive matrices $\boldsymbol{\phi}$ by the moving average $\boldsymbol{\theta}$ and the order p by q. Consider the mq-variate random vectors $\epsilon_t = (\boldsymbol{\varepsilon}'_t, \boldsymbol{\varepsilon}'_{t-1}, ..., \boldsymbol{\varepsilon}'_{t-q+1})'$, $\boldsymbol{\upsilon}_t = (\mathbf{y}'_t, \mathbf{0}_{1\times m(q-1)})'$, and the $S-$periodic matrix $\boldsymbol{\Phi}_t$ of dimension $mp \times mp$:

$$\boldsymbol{\Theta}_t = \left(\begin{array}{c|c} \boldsymbol{\theta}_{t,1} \ldots \boldsymbol{\theta}_{t,q-1} & \boldsymbol{\theta}_{t,q} \\ \hline \mathbf{I}_{m(q-1)\times m(q-1)} & \mathbf{0}_{m(q-1)\times m} \end{array} \right),$$

where $q = \max_{1\leq s\leq S}\{q_s\}$, then one can rewrite model (1.2.3) in the following equivalent mq-variate S-periodic $PVMA_S(1)$ model

$$\epsilon_t = \boldsymbol{\Theta}_t \epsilon_{t-1} + \boldsymbol{\upsilon}_t. \tag{1.2.17}$$

Let $\boldsymbol{\Theta} = \prod_{s=1}^{S} \boldsymbol{\Theta}_{S-s+1}$. Then the following proposition gives a necessary and sufficient invertibility condition for (1.2.11) and hence (1.2.3).

Proposition 1.2.3

The $m-$variate, S-periodic $PVMA$ model (1.2.3) *is invertible if and only if*

$$\rho\left(\boldsymbol{\Theta}\right) < 1. \tag{1.2.18}$$

Proof

The proof is similar to that of Proposition 1.2.2 and hence it will be omitted. ■

This condition seems much simpler than the other existing conditions in he literature (Tiao and Grupe, 1980; Bentarzi and Hallin, 1994).

1.3 The autocorrelation structure

Despite the importance of periodic $VARMA$ autocovariances for statistical inference purposes, their calculation has received little attention. It is worth noting that the autocovariances calculation is more tedious for periodic $ARMA$ models than for their stationary counterparts, mainly because of the absence of symmetry and hence the Toeplitz property in the autocovariance structure. Indeed, it appears that the periodic autocovariance function cannot be computed by recursive methods such as the most efficient algorithm of Wilson (1979), which is devoted exclusively to stationary univariate $ARMA$ autocovariances. Based on the difference equation approach introduced by McLeod (1975) for the standard $ARMA$ case, Shao and Lund (2004) and Li and Hui (1988) have studied the autocovariance structure of univariate $PARMA_S(p,q)$ models. Furthermore, Boteva and Boshnakov (1992) and Bentarzi and Aknouche (2005) considered the $PAR_S(p)$ autocovariance calculation. A serious drawback of this approach is the need to solve a linear system of order $(p+1)S$, from which the initial autocovariances (i.e., those of lag less than p) of all the seasons are computed in the same block. This makes the calculation very costly for models with a large period. Based on a state space representation of $PVAR$ models, we derive an efficient procedure to compute the autocovariance function of causal $PVAR$ models, where the autocovariance function for a fixed season, say s, is calculated and from which the autocovariances corresponding to the other seasons are easily deduced. Based on a state space representation of $PVARMA$ models, we derive an efficient algorithm for computing the autocovariance functions of causal $PVARMA$ models, where the autocovariances for a fixed season, say s, are calculated and from where the autocovariances corresponding to the other seasons are easily deduced. To simplify the exposition we present in the first subsection autocovariance calculations for $PVAR$ models. The more general $PVARMA$ case will be considered in the second subsection.

1.3.1 Calculation of the $PVAR$ autocovariances

This subsection deals with the calculation of theoretical autocovariances of an $m-$variate $PVAR$ model by exploiting the periodic state space representation (1.2.13). The technique used here will be generalized below to deal with a general $m-$variate $PVARMA$ model. Letting $t = s + nS$, where $n = \left\lceil \frac{t-1}{S} \right\rceil$, $s = 1, ..., S$ and t, $n \in \mathbb{Z}$, the periodic model (1.2.13) can be rewritten in the form

$$\mathbf{Y}_{s+nS} = \mathbf{\Phi}_s \mathbf{Y}_{s-1+nS} + \boldsymbol{\eta}_{s+nS}, \ \ s = 1, ..., S, \, n \in \mathbb{Z}, \tag{1.3.1}$$

where $E\left(\boldsymbol{\eta}_{s+nS}\boldsymbol{\eta}'_{s-h+nS}\right) = \delta_{h0}\Delta_s$, $\Delta_s = \begin{pmatrix} \Lambda_s & \mathbf{0}_{m\times m(p-1)} \\ \mathbf{0}_{m(p-1)\times m} & \mathbf{0}_{m(p-1)\times m(p-1)} \end{pmatrix}$, and $\Lambda_s = E\left(\boldsymbol{\varepsilon}_{s+nS}\right.$ $\left.\boldsymbol{\varepsilon}'_{s+nS}\right)$. Let $\boldsymbol{\gamma}_h^{(t)} = E\left(\mathbf{y}_t\mathbf{y}'_{t-h}\right)$ and $\Gamma_h^{(t)} = E\left(\mathbf{Y}_t\mathbf{Y}'_{t-h}\right)$ be the autocovariance functions at time t and lag h of $\{\mathbf{y}_t, t \in \mathbb{Z}\}$ and $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$, respectively. Then $\boldsymbol{\gamma}_h^{(t)}$ and $\Gamma_h^{(t)}$ are S-periodic in t, i.e., $\boldsymbol{\gamma}_h^{(s+nS)} = \boldsymbol{\gamma}_h^{(s)}$, $\Gamma_h^{(s+nS)} = \Gamma_h^{(s)}$, $s = 1, ..., S$, and we have

$$\Gamma_h^{(s)} = \begin{pmatrix} \boldsymbol{\gamma}_h^{(s)} & \boldsymbol{\gamma}_{h+1}^{(s)} & \cdots & \boldsymbol{\gamma}_{h+p-1}^{(s)} \\ \boldsymbol{\gamma}_{h-1}^{(s-1)} & \boldsymbol{\gamma}_h^{(s-1)} & \cdots & \boldsymbol{\gamma}_{h+p-2}^{(s-1)} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\gamma}_{h-p+1}^{(s-p+1)} & \boldsymbol{\gamma}_{h-p+2}^{(s-p+1)} & \cdots & \boldsymbol{\gamma}_h^{(s-p+1)} \end{pmatrix},$$

with the property $\boldsymbol{\gamma}_{-h}^{(s)} = \left(\boldsymbol{\gamma}_{h}^{(s+h)}\right)'$.

It is possible to obtain the $PVAR$ autocovariances via a difference equation approach (e.g. Li and Hui, 1988 and Shao and Lund, 2004; Bentarzi and Aknouche 2005), which calls for solving a block system, obtained from the periodic Yule-Walker equations (see Pagano, 1978 and Vecchia 1985b for the univariate case). As is well-known, from the autocovariance function these equations allow the calculation of the $PVAR_S(p)$ parameters by solving S independent linear systems of order mp, for each season.

Indeed, postmultiplying both sides of (1.3.1) by $\mathbf{y}'_{t-h}$, $h = 0, ..., p$, and then taking expectation, we obtain S linear systems of equations (of order pm) of the form

$$\begin{cases} \boldsymbol{\phi}_s \Gamma_s \ = \underline{\boldsymbol{\gamma}}_s \\ \Lambda_s = \boldsymbol{\gamma}_0^{(s)} - \boldsymbol{\phi}_s \underline{\boldsymbol{\gamma}}_s \end{cases}, \ s = 1, ..., S, \tag{1.3.2a}$$

where $\boldsymbol{\phi}_s = \left(\boldsymbol{\phi}_{s,1}, \boldsymbol{\phi}_{s,2}, ..., \boldsymbol{\phi}_{s,p}\right)$, $(\Gamma_s)_{kj} = \boldsymbol{\gamma}_{k-j}^{(s-j)}$, k, $j = 1, ..., p$ and $\underline{\boldsymbol{\gamma}}_s = (\boldsymbol{\gamma}_1^{(s)}, \boldsymbol{\gamma}_2^{(s)}, \ ..., \boldsymbol{\gamma}_p^{(s)})$.

The reverse problem which consists of finding the autocovariances $\boldsymbol{\gamma}_h^{(s)}$ from given model parameters can be formulated by rearranging (1.3.2a) to obtain the following system, which we solve with respect to $\underline{\gamma}$

$$G\, \underline{\gamma} = \underline{g}, \tag{1.3.2b}$$

where $\underline{\gamma} = (\gamma_0^{(1)\prime}, ..., \gamma_p^{(1)\prime}, \gamma_0^{(2)\prime}, ..., \gamma_p^{(2)\prime}, ..., \gamma_0^{(S)\prime}, ..., \gamma_p^{(S)\prime})'$ is the matrix of initial autocovariances, G is an $mS\,(p+1)$-square circular matrix of some suitable elements and $\underline{g}$ is an appropriate matrix whose form is unimportant here. So, the calculation of the initial autocovariances requires the resolution of a block system (of order $mS(p+1)$) associated to all the seasons. This makes the calculation very costly for models with a large period. Moreover, explicit expressions for G seem very difficult to obtain in the multivariate case. The following example clarifies this approach.

Example

Consider the following second order 2-periodic $PAR_2(2)$ model

$$y_{i+2\tau} - \phi_{i1} y_{i+2\tau-1} - \phi_{i2} y_{i+2\tau-2} = \varepsilon_{i+2\tau}, \quad i = 1, 2, \ \tau \in \mathbb{Z},$$

where $\{\varepsilon_t, \ t \in \mathbb{Z}\}$ is a periodic white noise with variance σ_t^2. Then, the periodic Yule-Walker equations are given by

$$\begin{cases} \gamma_1^{(1)}\phi_{11} + \gamma_2^{(1)}\phi_{12} = \gamma_0^{(1)} - \sigma_1^2 \\ \gamma_0^{(2)}\phi_{11} + \gamma_1^{(2)}\phi_{12} = \gamma_1^{(1)} \\ \gamma_1^{(2)}\phi_{11} + \gamma_0^{(1)}\phi_{12} = \gamma_2^{(1)} \end{cases}$$
$$\begin{cases} \gamma_1^{(2)}\phi_{21} + \gamma_2^{(2)}\phi_{22} = \gamma_0^{(2)} - \sigma_2^2 \\ \gamma_1^{(1)}\phi_{21} + \gamma_1^{(1)}\phi_{12} = \gamma_1^{(2)} \\ \gamma_1^{(1)}\phi_{21} + \gamma_0^{(2)}\phi_{22} = \gamma_2^{(2)} \end{cases} \tag{1.3.3}$$

where $\gamma_h^{(i)}$, is the autocovariance of season i and lag h. To calculate the initial autocovariances $\gamma_h^{(i)}$, $i = 1, 2$ and $h = 0, 1$; one can transform (1.3.3) to the following system

$$\left[\begin{array}{ccc|ccc} 1 & -\phi_{11} & -\phi_{12} & 0 & 0 & 0 \\ 0 & 1 & 0 & -\phi_{11} & -\phi_{12} & 0 \\ -\phi_{12} & 0 & 1 & 0 & -\phi_{11} & 0 \\ \hline 0 & 0 & 0 & 1 & -\phi_{21} & -\phi_{22} \\ -\phi_{21} & -\phi_{22} & 0 & 0 & 1 & 0 \\ 0 & -\phi_{21} & 0 & -\phi_{22} & 0 & 1 \end{array}\right] \left[\begin{array}{c} \gamma_0^{(1)} \\ \gamma_1^{(1)} \\ \gamma_2^{(1)} \\ \hline \gamma_0^{(2)} \\ \gamma_1^{(2)} \\ \gamma_2^{(2)} \end{array}\right] = \left[\begin{array}{c} \sigma_1^2 \\ 0 \\ 0 \\ \hline \sigma_2^2 \\ 0 \\ 0 \end{array}\right],$$

which we solve with respect to $\gamma_h^{(i)}$.

Another well-known approach for calculating the autocovariances $\boldsymbol{\gamma}_h^{(s)}$ uses the causal representation (1.2.16) of (1.2.2) (see Lund and Basawa, 1999 and 2000 for the univariate $PARMA$ case). The autocovariances $\boldsymbol{\gamma}_h^{(s)}$ can be obtained, for all $s = 1, ..., S$ and $h \in \mathbb{Z}$, by

$$\boldsymbol{\gamma}_h^{(s)} = \sum_{j=0}^{+\infty} \boldsymbol{\psi}_{s,j} \Lambda_{s-j} \boldsymbol{\psi}'_{s-h,h+j},$$

where the previous infinite sum will be truncated under the causality condition, taking into account the geometric convergence of $\boldsymbol{\psi}_{s,k}\Lambda_{s-k}\boldsymbol{\psi}'_{s-h,h+k}$ to zero as $k \to \infty$. However, the computation of this latter expression is very tedious, particularly for large dimensions and when the parameters lie near the boundary of the causal domain.

Based on the state space representation (1.3.1), we propose another method, which computes the $PVAR$ autocovariances for distinct seasons separately, thereby facilitating efficient computation.

Let $\otimes$ be the Kronecker product, $vec(A)$ the usual column stacking and $A^{\otimes m}$ the Kronecker product $A \otimes A \otimes \cdots \otimes A$ of m factors. Using these notations, we can state the following proposition, which establishes a simple computational expression of the theoretical autocovariances, while avoiding the above inconveniences.

Proposition 1.3.1 (Aknouche, 2006)

For a causal *PVAR model, the autocovariances are given by the following equations*

$$vec\left(\Gamma_0^{(s)}\right) = \left(\mathbf{I}_{(mp)^2\times(mp)^2} - \prod_{k=1}^{S}\mathbf{\Phi}_{s-k+1}^{\otimes 2}\right)^{-1}\left(vec\left(\Delta_s\right) + \sum_{j=1}^{S-1}\prod_{k=1}^{j}\mathbf{\Phi}_{s-k+1}^{\otimes 2} vec\left(\Delta_{s-j}\right)\right), \quad (1.3.4a)$$

and

$$vec\left(\Gamma_h^{(s)}\right) = \prod_{j=1}^{h}\left(I_{mp\times mp}\otimes\mathbf{\Phi}_{s-j+1}\right) vec\left(\Gamma_0^{(s-h)}\right), \ h\in\mathbb{N}^*;\ s=1,2,...,S. \quad (1.3.4b)$$

Proof

Suppose that (1.2.14) holds. Postmultiplying both sides of (1.3.1) by $\mathbf{Y}'_{s-h+nS}$, where h is a positive integer, and then taking expectation, we find the periodic Yule-Walker equations

$$\Gamma_0^{(s)} = \mathbf{\Phi}_s\Gamma_1^{(s)\prime} + \Delta_s, \quad (1.3.5a)$$

$$\Gamma_h^{(s)} = \mathbf{\Phi}_s\Gamma_{h-1}^{(s-1)}, \qquad h\in\mathbb{N}^*;\ s=1,2,...,S. \quad (1.3.5b)$$

From $(1.3.5a)$ and $(1.3.5b)$ with $h=1$, we obtain the following *discrete-time periodic Lyapunov equation* (e.g., Bittanti et al., 1988; Varga 1997)

$$\Gamma_0^{(s)} = \mathbf{\Phi}_s\Gamma_0^{(s-1)}\mathbf{\Phi}'_s + \Delta_s. \quad (1.3.5c)$$

Taking the column stacking operator of both sides of $(1.3.5c)$ and applying the property $vec(ABC) = (C'\otimes A)\, vec(B)$, we find the first order linear difference equation

$$vec\left(\Gamma_0^{(s)}\right) = \mathbf{\Phi}_s^{\otimes 2} vec\left(\Gamma_0^{(s-1)}\right) + vec\left(\Delta_s\right). \quad (1.3.6)$$

After r $(r\geq 1)$ successive replacements in the latter equation, we obtain

$$vec\left(\Gamma_0^{(s)}\right) = \prod_{k=1}^{r}\mathbf{\Phi}_{s-k+1}^{\otimes 2} vec\left(\Gamma_0^{(s-r)}\right) + vec\left(\Delta_s\right) + \sum_{j=1}^{r-1}\prod_{k=1}^{j}\mathbf{\Phi}_{s-k+1}^{\otimes 2} vec\left(\Delta_{s-j}\right). \quad (1.3.7)$$

Putting $r = S$ in (1.3.7) and taking into account the S-periodicity of $\Gamma_0^{(s)}$, we get (1.3.4a). Notice that the invertibility of $\left(\mathbf{I}_{(mp)^2\times(mp)^2} - \prod_{k=1}^{S} \mathbf{\Phi}_{s-k+1}^{\otimes 2}\right)$ follows from the periodic stationarity of model (1.3.1). Indeed, the eigenvalues of $\prod_{k=1}^{S} \mathbf{\Phi}_{s-k+1}^{\otimes 2} = \left(\prod_{k=1}^{S} \mathbf{\Phi}_{s-k+1}\right)^{\otimes 2}$ are the product of the eigenvalues of $\prod_{k=1}^{S} \mathbf{\Phi}_{s-k+1}$ and hence (by a cyclical rearrangement) of $\mathbf{\Omega}$ in (1.2.14).

Otherwise, using the property $vec(AB) = (\mathbf{I} \otimes A)\, vec(B)$, equation (1.3.5$b$) can be written in a stacking column form as

$$vec\left(\Gamma_h^{(s)}\right) = \left(\mathbf{I}_{mp\times mp} \otimes \mathbf{\Phi}_s\right) vec\left(\Gamma_{h-1}^{(s-1)}\right). \tag{1.3.8}$$

Finally, by successive substitution in (1.3.8) we get (1.3.4b). ■

Though (1.3.4a) and (1.3.4b) express the autocovariances in a simple form, they have however two main drawbacks. First, for each season $s \in \{1, ..., S\}$ we need to invert a matrix of order $mp \times mp$, which is computationally cumbersome for large orders and/or periods. Second, the vector $vec\left(\Gamma_0^{(s)}\right)$ contains some identical entries because of the symmetry and the periodicity (this occurs only when $p > S$) of $\Gamma_0^{(s)}$, thereby causing an undesirable redundancy in the calculation. The following corollary allows one to avoid the first drawback, while the second one is remedied below.

Corollary 1.3.1 (Aknouche, 2006)

Expressions (1.3.4a) and (1.3.4b) *are equivalent to*

$$\left(\mathbf{I}_{(mp)^2} - \prod_{k=1}^{S} \mathbf{\Phi}_{S-k+1}^{\otimes 2}\right) vec\left(\Gamma_0^{(S)}\right) = \left(vec\left(\Lambda_S\right) + \sum_{j=1}^{S-1} \prod_{k=1}^{S} \mathbf{\Phi}_{S-k+1}^{\otimes 2} vec\left(\Lambda_{S-j}\right)\right), \tag{1.3.9}$$

$$vec\left(\Gamma_0^{(s)}\right) = \prod_{k=1}^{s} \mathbf{\Phi}_{s-k+1}^{\otimes 2} vec\left(\Gamma_0^{(S)}\right) + vec\left(\Lambda_s\right) + \sum_{j=1}^{s-1} \prod_{k=1}^{j} \mathbf{\Phi}_{s-k+1}^{\otimes 2} vec\left(\Lambda_{s-j}\right), \tag{1.3.10}$$

and

$$\Gamma_h^{(s)} = \mathbf{\Phi}_s \Gamma_{h-1}^{(s-1)}.$$

Furthermore, the block geometric structure of the autocovariances is characterized by

$$\Gamma_{\nu+\tau S}^{(s)} = \left(\prod_{j=1}^{S} \mathbf{\Phi}_{s-j+1}\right)^{\tau} \Gamma_{\nu}^{(s)}, \qquad \nu,\ s = 1, ..., S,\ \tau \in \mathbb{N}. \tag{1.3.11}$$

Proof

Putting $r = s$ in (1.3.7) we get (1.3.10) and replacing s by S in (1.3.10) we find (1.3.9). Expressing the lag h in the form $h = \nu + nS$, $\nu = 1, ..., S$ and $n \in \mathbb{N}$, one rewrites (1.3.4a) in the form

$$\begin{aligned}
\Gamma_{h}^{(s)} &= \mathbf{\Phi}_s \Gamma_{h-1}^{(s-1)} = \mathbf{\Phi}_s\, \mathbf{\Phi}_{s-1} \ldots \mathbf{\Phi}_{s-h+1}\, \Gamma_0^{(s-h)}, \quad h \in \mathbb{N}^*, \\
&= \left(\mathbf{\Phi}_s\, \mathbf{\Phi}_{s-1} \ldots \mathbf{\Phi}_{s-S+1}\right)^n \mathbf{\Phi}_s\, \mathbf{\Phi}_{s-1} \ldots \mathbf{\Phi}_{s-\nu+1}\, \Gamma_0^{(s-\nu)}, \\
&= \left(\prod_{j=1}^{S} \mathbf{\Phi}_{s-j+1}\right)^{n} \Gamma_{\nu}^{(s)},\ h = \nu + nS, \qquad v, s = 1, 2, ..., S.
\end{aligned}$$

■

From (1.3.11) we can see that the autocorrelation functions of a periodic VAR are characterized by a geometric decay by a block of S lags. This feature is useful for the model identification stage. It is worth noting that the computation of $\Gamma_0^{(s)}$, $s = 1, ..., S$, via (1.3.9)-(1.3.11) need only solve one linear system of order m^2p^2 given by (1.3.9). Moreover, this computation can be improved as follows.

i) For the case where $S > p$, which is very often encountered in periodic time series models, we know that the redundancy in $\Gamma_0^{(S)}$ is caused only by its symmetry. Hence we can use the half colon stacking operator $vech$ rather than vec, by means of the formula $vec(A) = D_{k\times k} vech(A)$, where A is a symmetric matrix of order k and $D_{k\times k}$ is the usual duplication matrix (e.g., Magnus and Neudecker, 1989). Thus, equation (1.3.9) can be rewritten in the equivalent form

$$\mathbf{B} vech\left(\Gamma_0^{(S)}\right) = \mathbf{C}, \tag{1.3.12}$$

where the square matrix $\mathbf{B}$ of dimension $\frac{pm(pm+1)}{2}$ is the first block of

$$\left(\mathbf{I}_{(mp)^2\times(mp)^2} - \prod_{k=1}^{S} \mathbf{\Phi}_{S-k+1}^{\otimes 2}\right) D_{(mp)^2\times(mp)^2},$$

and $\mathbf{C}$ is the vector containing the first $\frac{pm(pm+1)}{2}$ entries of the right-hand side of (1.3.9). Consequently, we need only solve one linear system of order $\frac{pm(pm+1)}{2}$. This reduction is, of course, insignificant for large models (dimension and/or order) but can be important when the autocovariance computation must be repeated many times, such as with maximum likelihood estimation.

ii) For the case $S \leq p$, it is possible to improve the computational complexity of (1.3.9) by stacking all the distinct elements of $\Gamma_0^{(S)}$ and using an appropriate permutation matrix rather than the duplication matrix $D_{k\times k}$ to formulate a similar linear system to (1.3.12).

Remark 1.3.1

The main difficulty involved in the autocovariance calculation when using Proposition 1.3.1 lies in solving *a discrete-time Lyapunov equation*

$$\Gamma = \mathbf{\Phi}\Gamma\mathbf{\Phi}' + \Delta,$$

for Γ. This equation has been widely studied in the engineering literature (see, for example, Barnett, (1974); Barraud (1977); Kitagawa, (1977) and Anderson and Moore, (1979, p. 67) for the case of unspecified solutions and Hammarling, (1982) for the case of nondefinite definite solutions). However, there are two additional complexities caused by the periodicity in our case. First, to obtain the above Lyapunov equation (see also (1.3.4a)), we need to form explicitly some matrix products and sums of matrix products that will increase the computational burden. Second, for a $PVAR_S(p)$ model with $p > S$, the matrix $\Gamma^{(s)}$ contains some redundant entries because of the periodicity of $\boldsymbol{\gamma}_h^{(t)}$ (see the examples below).

The main usefulness of the proposed method given by (1.3.4) or equivalently by (1.3.9)-(1.3.12) is that the autocovariance function of each season is computed separately, contrary to the existing methods in the literature (cf. Li and Hui, 1988; Shao and Lund, 2004; Bentarzi and Aknouche, 2005). The following examples illustrate the usefulness of formulae (1.3.4), (1.3.9), and (1.3.10).

Examples

i) Consider the first order m-variate S-periodic autoregressive model ($PVAR_S(1)$)

$$\mathbf{y}_t = \boldsymbol{\phi}_t \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t,\ t \in \mathbb{Z}, \tag{1.3.13a}$$

which has the state space representation in (1.3.1). Then, $\Gamma_h^{(s)} = \boldsymbol{\gamma}_h^{(s)}$, $\mathbf{\Phi}_t = \boldsymbol{\phi}_t$ and $\Delta_t = \Lambda_t$. Using the difference equation approach (1.3.2b), the calculation of the initial $PVAR_S(1)$ autocovariances requires the resolution of a system of order $2mS$, which is very tedious, for instance, for monthly data with period 12. However, from Corollary 1.3.1, this calculation can be achieved more easily.

Applying (1.3.9), (1.3.10), and assuming causality, we obtain the initial autocovariances directly by solving a system of order m^2 :

$$\left(\mathbf{I}_{m^2\times m^2}-\prod_{k=1}^{S}\boldsymbol{\phi}_{S-k+1}^{\otimes 2}\right) vec\left(\boldsymbol{\gamma}_0^{(S)}\right)=\left(vec\left(\Lambda_S\right)+\sum_{j=1}^{S-1}\prod_{k=1}^{S}\boldsymbol{\phi}_{S-k+1}^{\otimes 2} vec\left(\Lambda_{S-j}\right)\right) \quad (1.3.14a)$$

$$vec\left(\boldsymbol{\gamma}_0^{(s)}\right)=\prod_{k=1}^{s}\boldsymbol{\phi}_{s-k+1}^{\otimes 2} vec\left(\boldsymbol{\gamma}_0^{(S)}\right)+vec\left(\Lambda_s\right)+\sum_{j=1}^{s-1}\prod_{k=1}^{j}\boldsymbol{\phi}_{s-k+1}^{\otimes 2} vec\left(\Lambda_{s-j}\right), \quad (1.3.14b)$$

and

$$\boldsymbol{\gamma}_h^{(s)}=\prod_{j=1}^{h}\boldsymbol{\phi}_{s-j+1}\boldsymbol{\gamma}_0^{(s-h)},\ \ s=1,2,...,S \text{ and } h\in\mathbb{N}. \quad (1.3.14c)$$

For the particular $PAR_S(1)$ model (i.e. when $m=1$ in (1.3.13a))

$$y_t=\phi_t y_{t-1}+\varepsilon_t,\ \ t\in\mathbb{Z}, \quad (1.3.13b)$$

where $\{\varepsilon_t,\ t\in\mathbb{Z}\}$ is a periodic white noise with variance σ_t^2, expressions (1.3.14a) and (1.3.14b) reduce to

$$\gamma_0^{(s)}=\left(1-\prod_{k=1}^{S}\phi_{s-k+1}^2\right)^{-1}\left(\sigma_s^2+\sum_{j=1}^{S-1}\prod_{k=1}^{j}\phi_{s-k+1}^2\sigma_{s-j}^2\right),\ \ s=1,...,S, \quad (1.3.14d)$$

and

$$\gamma_h^{(s)}=\prod_{j=1}^{h}\phi_{s-j+1}\,\gamma_0^{(s-h)},\ \ s=1,2,...,S \textit{ and } h\in\mathbb{N}, \quad (1.3.14e)$$

which have previously been obtained by Lund and Basawa (1999, Example 4.2). From Corollary 1.3.1, one can express the variances $\gamma_0^{(s)}$ in the equivalent form

$$\begin{aligned}\gamma_0^{(S)} &= \left(1-\prod_{k=1}^{S}\phi_k^2\right)^{-1}\left(\sigma_S^2+\sum_{j=1}^{S-1}\prod_{k=1}^{s}\phi_{S-k+1}^2\,\sigma_{S-j}^2\right),\\ \gamma_0^{(s)} &= \gamma_0^{(S)}\prod_{k=1}^{j}\phi_{s-k+1}^2+\sigma_s^2+\sum_{j=1}^{s-1}\prod_{k=1}^{j}\phi_{s-k+1}^2\,\sigma_{s-j}^2,\end{aligned}$$

Note that expression (1.3.14e) can be exploited to reveal the periodic autocorrelation structure of the processes. Indeed, dividing both sides of (1.3.14e) by $\sqrt{\gamma_0^{(s)}}\sqrt{\gamma_0^{(s-h)}}$, we obtain the autocorrelation at lag h for the $s-th$ season, $\rho_h^{(s)}$

$$\rho_h^{(s)} = \prod_{j=1}^{h} \phi_{s-j+1} \frac{\sqrt{\gamma_0^{(s-h)}}}{\sqrt{\gamma_0^{(s)}}},\ h \in \mathbb{N},\ s = 1, 2, ..., S.$$

Writing the lag h in the form $h = \nu + nS$, $\nu = 1, ..., S$ and $n \in \mathbb{N}$, the latter expression becomes

$$\begin{aligned} \rho_{\nu+nS}^{(s)} &= \prod_{j=1}^{\nu+nS} \phi_{s-j+1} \frac{\sqrt{\gamma_0^{(s-\nu+nS)}}}{\sqrt{\gamma_0^{(s)}}} = \left(\prod_{j=1}^{S} \phi_j\right)^n \prod_{j=1}^{\nu} \phi_{s-j+1} \frac{\sqrt{\gamma_0^{(s-\nu)}}}{\sqrt{\gamma_0^{(s)}}} \\ &= \varphi^n \rho_\nu^{(s)}, \end{aligned}$$

where $\varphi = \prod_{j=1}^{S} \phi_j$, $v, s \in \{1, ..., S\}$ *and* $n \in \mathbb{N}$. The latter equation clearly shows the S-block geometric decay of the autocorrelations as mentioned in Corollary 1.3.1. For instance, taking $S = 4$ in (1.3.13b) with $\phi = [\phi_1, \phi_2, \phi_3, \phi_4] = [0.9,\ 0.85,\ 0.66,\ 0.45]$ and $\sigma^2 = \left[\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2\right] = [1,\ 1,\ 1,\ 1]'$, the autocorrelation functions $\rho_h^{(s)}$, $s = 1, ..., 4$, are represented by Figure 1.

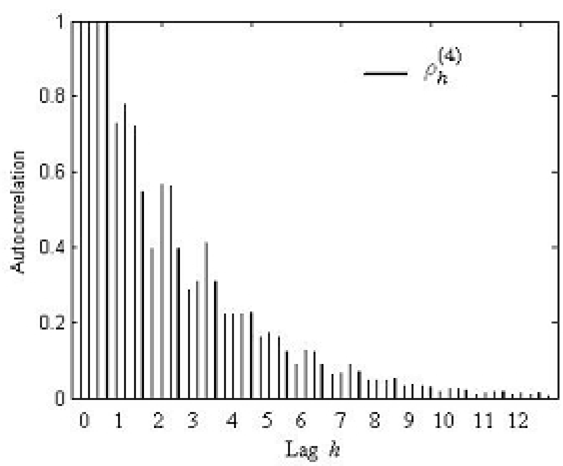


Figure 1 : $PAR_4(1)$ autocorrelations

ii) Consider the m-variate 2-periodic autoregressive model of order 4 ($PVAR_2(4)$)

$$\mathbf{y}_t = \boldsymbol{\phi}_{t,1}\mathbf{y}_{t-1} + \boldsymbol{\phi}_{t,2}\mathbf{y}_{t-2} + \boldsymbol{\phi}_{t,3}\mathbf{y}_{t-3} + \boldsymbol{\phi}_{t,4}\mathbf{y}_{t-4} + \boldsymbol{\varepsilon}_t.$$

Since in this case the period is lower than the order, we can see that the matrix

$$\Gamma_0^{(S)} = \begin{pmatrix} \gamma_0^{(S)} & \gamma_1^{(S)} & \gamma_2^{(S)} & \gamma_3^{(S)} \\ \gamma_1^{(S)\prime} & \gamma_0^{(S-1)} & \gamma_1^{(S-1)} & \gamma_2^{(S-1)} \\ \gamma_2^{(S)\prime} & \gamma_1^{(S-1)\prime} & \gamma_0^{(S)} & \gamma_1^{(S)} \\ \gamma_3^{(S)\prime} & \gamma_2^{(S-1)\prime} & \gamma_1^{(S)\prime} & \gamma_0^{(S-1)} \end{pmatrix}$$

contains some identical entries and hence its computation by means of Corollary 1.3.1 seems cumbersome. However, as emphasized above, by stacking all the distinct elements of $\Gamma_0^{(S)}$ and using an appropriate permutation matrix rather than the duplication matrix $D_{k\times k}$, we can formulate a similar linear system to (1.3.12), which computes the autocovariances $\gamma_h^{(s)}$ without any computational redundancy. This example clearly shows one of the difficulties caused by the periodic case, as mentioned in Remark 1.3.1.

1.3.2 Calculation of the $PVARMA$ autocovariances

In what follows we extend the previous autocovariance calculation algorithm to the general $PVARMA$ case. With $t = s + S\tau$ rewriting model (1.2.1) as follows

$$\mathbf{y}_{s+S\tau} - \sum_{j=1}^{p} \boldsymbol{\phi}_{s,j}\, \mathbf{y}_{s-j+S\tau} = \boldsymbol{\varepsilon}_{s+S\tau} - \sum_{j=1}^{q} \boldsymbol{\theta}_{s,j} \boldsymbol{\varepsilon}_{s-j+S\tau},\ s = 1, ..., S,\ \ \tau = \left\lceil \frac{t-1}{S} \right\rceil \tag{1.3.15}$$

where $p = \max\limits_{i\in\{1,2,...,S\}} p_i$ and $q = \max\limits_{j\in\{1,2,...,S\}} q_j$. Under the causality assumption (1.2.14), the latter model admits the following causal representation

$$\mathbf{y}_t = \sum_{j=0}^{\infty} \boldsymbol{\psi}_{t,j} \boldsymbol{\varepsilon}_{t-j} \tag{1.3.16}$$

where the sequences $S-$periodic coefficients $\boldsymbol{\psi}_{t,j}$ are absolutely summable, in the sense that

$\sum_{j=0}^{+\infty} \left\| \boldsymbol{\psi}_{ij} \right\| < \infty$ for all t with "$\|.\|$" stands for the usual Euclidian norm. Like the previous $PVAR$ case, these coefficients can be obtained recursively in terms of the model parameters via the equations (see Lund and Basawa (2000) for the univariate $PARMA$ case)

$$\begin{cases} \boldsymbol{\psi}_{s,0} = I_m & \\ \boldsymbol{\psi}_{s,k} = -\boldsymbol{\theta}_{s,k} 1_{[k\leq q]} + \sum\limits_{j=1}^{\min(k,p)} \boldsymbol{\phi}_{s,j} \boldsymbol{\psi}_{s-j,k-j} & s = 1, ..., S. \end{cases} \tag{1.3.17}$$

Next, we exploit a state space representation of the $PVARMA(p,q)$ model (1.3.15) to obtain a simple form of the theoretical autocovariances.

As in the latter subsection define the $mp \times 1$ random vectors

$\mathbf{Y}_{s+nS} = (\mathbf{y}'_{s+nS}, \mathbf{y}'_{s+nS-1}..., \mathbf{y}'_{s+nS-p+1})'$, $\boldsymbol{\epsilon}_{s+nS} = (\boldsymbol{\xi}'_{s+nS}, \mathbf{0}_{1\times m(p-1)})'$ where

$$\boldsymbol{\xi}_{s+nS} = \boldsymbol{\varepsilon}_{s+nS} - \sum_{j=1}^{q} \boldsymbol{\theta}_{s,j}\boldsymbol{\varepsilon}_{s+nS-j}.$$

Using these notations, we can rewrite model (1.3.15) in the following equivalent periodic state space representation

$$\mathbf{Y}_{s+nS} = \boldsymbol{\Phi}_s \mathbf{Y}_{s+nS-1} + \boldsymbol{\epsilon}_{s+nS}, \tag{1.3.18}$$

where the $S-$periodic $(mp \times mp)$ matrix $\boldsymbol{\Phi}_t$ is defined as in Section 1.2.

let

$$\boldsymbol{\gamma}_h^{(s)} = E\left(\mathbf{y}_{s+nS}\mathbf{y}'_{s+nS-h}\right) \quad \text{and} \quad \boldsymbol{\Gamma}_h^{(s)} = E\left(\mathbf{Y}_{s+nS}\mathbf{Y}'_{s+nS-h}\right)$$

be the autocovariance functions at period s and lag $h \in \mathbb{Z}$, of $\{\mathbf{y}_t, t \in \mathbb{Z}\}$ and $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$, respectively. Let also $E\left(\boldsymbol{\epsilon}_{s+nS}\mathbf{Y}'_{s-h+nS}\right) = \Lambda_{s,h}$ be the cross autocovariance function at season s and lag $h \in \mathbb{Z}$. Then as in the former subsection, $\boldsymbol{\gamma}_h^{(t)}$, $\boldsymbol{\Gamma}_h^{(t)}$ and $\Lambda_{s,h}$ are S-periodic of t, (i.e., $\boldsymbol{\gamma}_h^{(s+nS)} = \boldsymbol{\gamma}_h^{(s)}$ and so on) and we have

$$\boldsymbol{\Gamma}_h^{(s)} = \begin{pmatrix} \boldsymbol{\gamma}_h^{(s)} & \boldsymbol{\gamma}_{h+1}^{(s)} & \cdots & \boldsymbol{\gamma}_{h+p-1}^{(s)} \\ \boldsymbol{\gamma}_{h-1}^{(s-1)} & \boldsymbol{\gamma}_h^{(s-1)} & \cdots & \boldsymbol{\gamma}_{h+p-2}^{(s-1)} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\gamma}_{h-p+1}^{(s-p+1)} & \boldsymbol{\gamma}_{h-p+2}^{(s-p+1)} & \cdots & \boldsymbol{\gamma}_h^{(s-p+1)} \end{pmatrix}.$$

Define for $h < q$ and $s = 1, ..., S$ the matrices $\boldsymbol{\Theta}_{s,h}$ and $\boldsymbol{\Psi}_{s,h}$ of dimensions $mp \times m(q-h+1)$ and $m(q-h+1) \times mp$, respectively, as

$$\boldsymbol{\Theta}_{s,h} = \begin{pmatrix} \boldsymbol{\theta}_{s,h}\boldsymbol{\Sigma}_{s-h} \quad \boldsymbol{\theta}_{s,h+1}\boldsymbol{\Sigma}_{s-h-1} \quad \cdots \quad \boldsymbol{\theta}_{s,q}\boldsymbol{\Sigma}_{s-q} \\ \mathbf{0}_{m(p-1)\times m(q-h+1)} \end{pmatrix}$$

and

$$\Psi_{s,h} = \begin{pmatrix} \mathbf{I}_m 1_{[q-h+1\geq 1]} & 0 & \cdots & 0 \\ \boldsymbol{\psi}'_{s-h,1} 1_{[q-h+1\geq 1]} & \mathbf{I}_m 1_{[q-h+1\geq 2]} & \cdots & \vdots \\ \vdots & \boldsymbol{\psi}'_{s-h-1,1} 1_{[q-h+1\geq 2]} & \cdots & 0 \\ \vdots & \vdots & \ddots & \mathbf{I}_m 1_{[q-h+1\geq p]} \\ \vdots & \ddots & \ddots & \vdots \\ \boldsymbol{\psi}'_{s-h,q-h} 1_{[q-h+1\geq 1]} & \boldsymbol{\psi}'_{s-h-1,q-h-1} 1_{[q-h+1\geq 2]} & \cdots & \boldsymbol{\psi}'_{s-h-p+1,q-h-p+1} 1_{[q-h+1\geq p]} \end{pmatrix}$$

where the $\boldsymbol{\psi}_{s,j}$ are the solutions of (1.3.17).

Using the previous formulations, the autocovariances of the $PVARMA$ model (1.3.15) can be obtained via the following proposition.

Proposition 1.3.2 (Aknouche and Hamdi, 2005)

For a causal $PVARMA$ model with $p \geq 1$, the autocovariances $\mathbf{\Gamma}_h^{(s)}$, for $1 \leq s \leq S$ and $h \in N$, are given by the following equations

$$\begin{aligned} \left(\mathbf{I}_{m^2p^2\times m^2p^2} - \prod_{k=1}^{S} \mathbf{\Phi}_{S-k+1}^{\otimes 2}\right) vec\left(\mathbf{\Gamma}_0^{(S)}\right) &= \sum_{j=0}^{S-1}\prod_{k=0}^{j-1} \mathbf{\Phi}_{S-k}^{\otimes 2}\left(\mathbf{I}_{mp\times mp}\otimes \mathbf{\Phi}_{S-j}\right) \\ &\times vec\left(\mathbf{\Lambda}'_{S-j,1}\right) + \sum_{j=0}^{S-1}\prod_{k=0}^{j-1} \mathbf{\Phi}_{S-k}^{\otimes 2} vec\left(\mathbf{\Lambda}_{S-j,0}\right), \end{aligned} \tag{1.3.19a}$$

$$\begin{aligned} vec\left(\mathbf{\Gamma}_0^{(s)}\right) = \prod_{j=1}^{s} \mathbf{\Phi}_{s-j+1}^{\otimes 2} vec\left(\mathbf{\Gamma}_0^{(S)}\right) &+ \sum_{j=0}^{s-1}\prod_{k=0}^{j-1} \mathbf{\Phi}_{s-k}^{\otimes 2}\left(\mathbf{I}_{mp\times mp}\otimes \mathbf{\Phi}_{s-j}\right) \\ \times vec\left(\mathbf{\Lambda}'_{s-j,1}\right) &+ \sum_{j=0}^{s-1}\prod_{k=0}^{j-1} \mathbf{\Phi}_{s-k}^{\otimes 2} vec\left(\mathbf{\Lambda}_{s-j,0}\right), \end{aligned} \tag{1.3.19b}$$

$$\mathbf{\Gamma}_h^{(s)} = \begin{cases} \prod_{k=0}^{h-1} \mathbf{\Phi}_{s-k}\mathbf{\Gamma}_0^{(s-h)} + \sum_{j=0}^{h-1}\prod_{k=0}^{j-1} \mathbf{\Phi}_{s-k}\mathbf{\Lambda}_{s-j,h-j}, & \text{for } 1 \leq h \leq q, \\ \prod_{k=0}^{h-1} \mathbf{\Phi}_{s-k}\mathbf{\Gamma}_0^{(s-h)}, & \text{for } h \geq q+1 \end{cases}, \tag{1.3.19c}$$

where

$$\mathbf{\Lambda}_{s,h} = -\mathbf{\Theta}_{s,h}\mathbf{\Psi}_{s,h}, \tag{1.3.20}$$

with the convention $\prod_{k=x}^{y} A_k = \mathbf{I}$ for $x > y$.

Proof

As for Proposition 1.3.1, postmultiplying both sides of (1.3.18) by $\mathbf{Y}'_{s-h+nS}$, $(h \in \mathbb{N}^*)$ and then taking expectations gives the following periodic Yule-Walker equations

$$\mathbf{\Gamma}_0^{(s)} = \mathbf{\Phi}_s \mathbf{\Gamma}_1^{(s)\prime} + \mathbf{\Lambda}_{s,0}, \tag{1.3.21a}$$

$$\mathbf{\Gamma}_h^{(s)} = \mathbf{\Phi}_s \mathbf{\Gamma}_{h-1}^{(s-1)} + \mathbf{\Lambda}_{s,h}, \qquad 1 \leq h \leq q \tag{1.3.21b}$$

$$\mathbf{\Gamma}_h^{(s)} = \mathbf{\Phi}_s \mathbf{\Gamma}_{h-1}^{(s-1)}, \qquad h \geq q+1, \tag{1.3.21c}$$

for $s = 1, ..., S$. From (1.3.21a) and (1.3.21b) with $h = 1$, we find the following *discrete-time periodic Lyapunov equation DTPLE* (e.g. Bittanti et al., 1988; Varga, 1997)

$$\mathbf{\Gamma}_0^{(s)} = \mathbf{\Phi}_s \mathbf{\Gamma}_0^{(s-1)} \mathbf{\Phi}'_s + \mathbf{\Phi}_s \mathbf{\Lambda}'_{s,1} + \mathbf{\Lambda}_{s,0},$$

and taking the column stacking operator of both sides of the above equation, we obtain the first-order linear difference equation

$$vec\left(\mathbf{\Gamma}_0^{(s)}\right) = \mathbf{\Phi}_s^{\otimes 2} vec\left(\mathbf{\Gamma}_0^{(s-1)}\right) + \left(\mathbf{I}_{mp\times mp} \otimes \mathbf{\Phi}_s\right) vec\left(\mathbf{\Lambda}'_{s,1}\right) + vec\left(\mathbf{\Lambda}_{s,0}\right),$$

where the property $vec(AB) = (\mathbf{I} \otimes A)\, vec(B)$ has been used. After r ($r \in \mathbb{N}^*$) successive replacements in the latter equation, we obtain

$$\begin{aligned} vec\left(\mathbf{\Gamma}_0^{(s)}\right) = {} & \prod_{j=1}^{r} \mathbf{\Phi}_{s-j+1}^{\otimes 2} vec\left(\mathbf{\Gamma}_0^{(s-r)}\right) + \sum_{j=0}^{r-1} \prod_{k=0}^{j-1} \mathbf{\Phi}_{s-k}^{\otimes 2} \left(\mathbf{I}_{mp\times mp} \otimes \mathbf{\Phi}_{s-j}\right) vec\left(\mathbf{\Lambda}'_{s-j,1}\right) \\ & + \sum_{j=0}^{r-1} \prod_{k=0}^{j-1} \mathbf{\Phi}_{s-k}^{\otimes 2} vec\left(\mathbf{\Lambda}_{s-j,0}\right). \end{aligned} \tag{1.3.22}$$

Taking now $r = s$ in (1.3.22), after some simple manipulations we get (1.3.19b). Replacing s by S in (1.3.19b) we find (1.3.19a). Finally, equations (1.3.19c) are obtained by successive substitution in (1.3.21b) and (1.3.21c), respectively. ■

Of course, for the case where $p = 0$, the autocovariance function of a pure periodic vector moving average ($PVMA$) model can be obtained from

$$\gamma_h^{(s)} = \begin{cases} \sum_{j=h}^{q} \boldsymbol{\theta}_{s,j} \Sigma_{s-j} \boldsymbol{\theta}'_{s-h,j-h} & \text{if} \quad 0 \leq h \leq q \\ \mathbf{0}_{m\times m} & \text{if} \quad h > q \end{cases}, \tag{1.3.23}$$

where $\boldsymbol{\theta}_{s,0} = \mathbf{I}_{m\times m}$, $s = 1, ..., S$ (see also Vecchia, (1985b) for the univariate $PARMA$ case).

Remark 1.3.2

As for Proposition 1.3.1, equation (1.3.19a) has the drawback of inverting a matrix of order $mp \times mp$. This seems computationally cumbersome for models with a large order or period. It is possible to reduce the computational complexity of the $PVARMA$ autocovariance calculation by using the same techniques as in (1.3.12).

It is worth mentioning that the main usefulness of the proposed method given by (1.3.19) is that the autocovariances for each season are computed separately, unlike existing methods in the literature (e.g. Li and Hui, 1988; Shao and Lund, 2004; Bentarzi and Aknouche, 2005). However, a possible drawback of our method lies in the need to form explicitly matrix products and sum of matrix products.

The following algorithm summarizes the computation of $PVARMA$ autocovariances.

Algorithm 1.3.1

For $p > 0$,

i) compute $\mathbf{\Lambda}_{s,h}$ from (1.3.20), where $\mathbf{\Psi}_{s,h}$ is obtained from (1.3.17);

ii) use equation (1.3.19a) to solve the following discrete-time periodic Lyapunov equation ($DTPLE$)

$$\mathbf{\Gamma}_0^{(s)} = \mathbf{\Phi}_s \mathbf{\Gamma}_0^{(s-1)} \mathbf{\Phi}_s' + \mathbf{\Phi}_s \mathbf{\Lambda}_{s,1}' + \mathbf{\Lambda}_{s,0}, \tag{1.3.24}$$

for $\Gamma_0^{(S)}$, and then deduce $\Gamma_0^{(s)}$, $s = 1, ..., S-1$ from (1.3.19b);

iii) for $h \geq 1$, calculate $\Gamma_h^{(s)}$, $s = 1, ..., S$ from (1.3.19c).

If $p = 0$ compute $\boldsymbol{\gamma}_h^{(s)}$ ($s = 1, ..., S$, $h \in \mathbb{N}^*$) from (1.3.23). ■

Remark 1.3.3

Instead of the vector stacking argument for solving equation (1.3.24) by (1.3.19a), the autocovariance $\mathbf{\Gamma}_0^{(S)}$ can be computed using two alternative approaches :

i) First, by reducing the $DTPLE$ (1.3.24) to a standard *discrete-time algebraic Lyapunov equation* ($DTALE$) of the form

$$\mathbf{\Gamma}_0^{(S)} = \mathbf{\Phi} \Gamma_0^{(S)} \mathbf{\Phi}' + \Delta,$$

and solving this one for $\mathbf{\Gamma}_0^{(S)}$ via standard exact methods (e.g. Barraud, 1977; Hammarling, 1982) or iterative methods (e.g. Anderson and Moore, 1979, p. 67).

ii) Second, by solving the *DTPLE* (1.3.24) directly (e.g. Sreedhar and Van Dooren 1994; Varga 1997) while avoiding the construction of matrix products via the *periodic Schur decomposition* (e.g., Bojanczyk et al., 1992).

Thought the methods of the second approach seem computationally superior, they do not, however, take into account the particular structure of $\mathbf{\Gamma}_0^{(S)}$.

We conclude this section with the following example, which illustrates the usefulness of Algorithm 1.3.1.

Example 3.5

Consider the m-variate S-periodic $PVARMA_S(1,q)$ model

$$\mathbf{y}_{s+nS} - \boldsymbol{\phi}_s\,\mathbf{y}_{s+nS-1} = \boldsymbol{\varepsilon}_{s+nS} - \sum_{j=1}^{q}\boldsymbol{\theta}_{s,j}\boldsymbol{\varepsilon}_{s-j+nS},\ \ s=1,...,S \text{ and } n\in\mathbb{Z},$$

which we rewrite in the equivalent periodic state space representation

$$\mathbf{y}_{s+nS} = \mathbf{\Phi}_s\mathbf{y}_{s+nS-1} + \boldsymbol{\epsilon}_{s+nS},$$

where $\boldsymbol{\epsilon}_{s+nS} = \boldsymbol{\varepsilon}_{s+nS} - \sum_{j=1}^{q}\boldsymbol{\theta}_{s,j}\boldsymbol{\varepsilon}_{s+nS-j}$, $\mathbf{\Phi}_s = \boldsymbol{\phi}_s$ and $\Gamma_h^{(s)} = \boldsymbol{\gamma}_h^{(s)}$. Using Proposition 1.3.2 under the causality and invertibility conditions, the $PVARMA_S(1,q)$ autocovariances are given by

$$\begin{aligned}\left(\mathbf{I}_{m^2\times m^2} - \prod_{k=1}^{S}\boldsymbol{\phi}_{S-k+1}^{\otimes 2}\right) vec\left(\boldsymbol{\gamma}_0^{(S)}\right) = \sum_{j=0}^{S-1}\prod_{k=0}^{j-1}\boldsymbol{\phi}_{S-k}^{\otimes 2}\left(\mathbf{I}_{m\times m}\otimes\boldsymbol{\phi}_{S-j}\right) vec\left(\mathbf{\Lambda}'_{S-j,1}\right)\\ + \sum_{j=0}^{S-1}\prod_{k=0}^{j-1}\boldsymbol{\phi}_{S-k}^{\otimes 2} vec\left(\mathbf{\Lambda}_{S-j,0}\right),\end{aligned} \tag{1.3.25a}$$

$$\begin{aligned}vec\left(\boldsymbol{\gamma}_0^{(s)}\right) = \prod_{j=1}^{s}\boldsymbol{\phi}_{s-j+1}^{\otimes 2} vec\left(\boldsymbol{\gamma}_0^{(S)}\right) + \sum_{j=0}^{s-1}\prod_{k=0}^{j-1}\boldsymbol{\phi}_{s-k}^{\otimes 2}\left(\mathbf{I}_{m\times m}\otimes\boldsymbol{\phi}_{s-j}\right) vec\left(\mathbf{\Lambda}'_{s-j,1}\right)\\ + \sum_{j=0}^{s-1}\prod_{k=0}^{j-1}\boldsymbol{\phi}_{s-k}^{\otimes 2} vec\left(\mathbf{\Lambda}_{s-j,0}\right),\end{aligned} \tag{1.3.25b}$$

and

$$\boldsymbol{\gamma}_h^{(s)} = \begin{cases}\boldsymbol{\phi}_s\boldsymbol{\gamma}_{h-1}^{(s-1)} + \mathbf{\Lambda}_{s,h}, & 1\le h\le q,\\ \boldsymbol{\phi}_s\boldsymbol{\gamma}_{h-1}^{(s-1)}, & h\ge q+1.\end{cases} \tag{1.3.25c}$$

Relations (1.3.25) reduce for $m = 1$ and $q = 1$ to the ones given by Lund and Basawa (2000, Example 5.1). For instance, for the particular bivariate 4-periodic $PVARMA_4(1,1)$ model (i.e. with $m = 2$, $S = 4$ and $p = q = 1$)

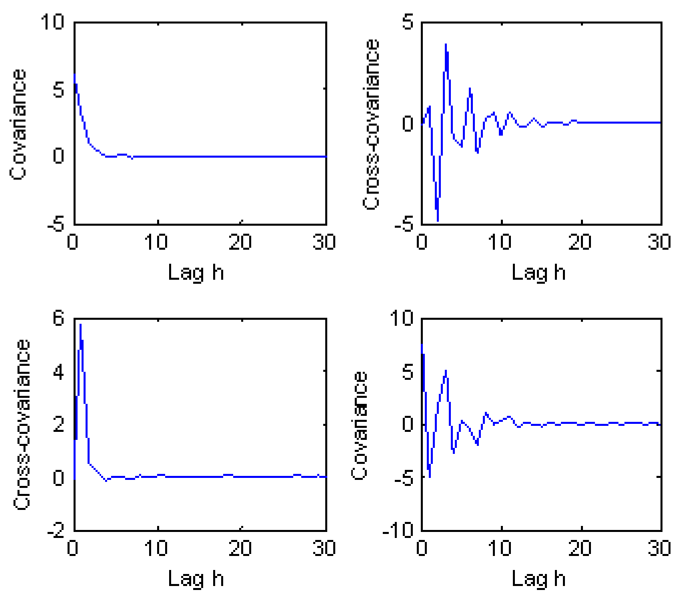
$$\mathbf{y}_{s+nS} - \boldsymbol{\phi}_s\, \mathbf{y}_{s+nS-1} = \boldsymbol{\varepsilon}_{s+nS} - \boldsymbol{\theta}_s \boldsymbol{\varepsilon}_{s+nS-1}, \;\; s = 1, ..., S \text{ and } n \in \mathbb{Z}, \qquad (1.3.26)$$

with parameters listed below

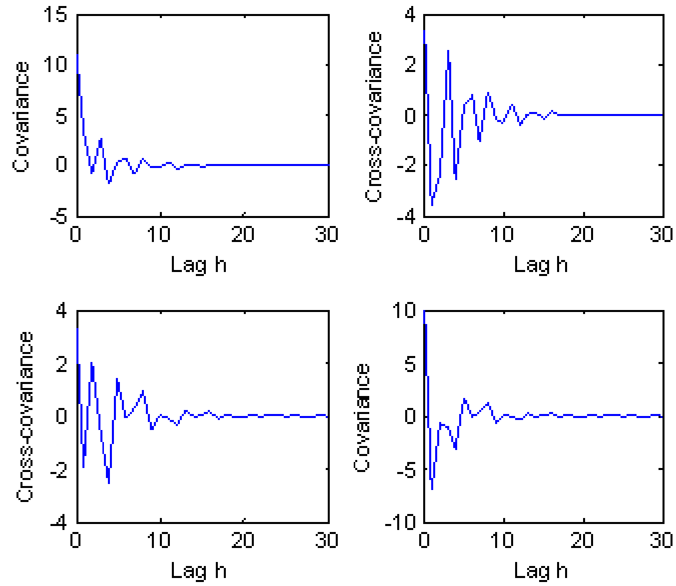
s	$\boldsymbol{\phi}_s$	$\boldsymbol{\theta}_s$	$\boldsymbol{\Sigma}_s$
1	$\begin{pmatrix} 0.30 & 0.50 \\ 0.45 & -0.20 \end{pmatrix}$	$\begin{pmatrix} 0.50 & -0.70 \\ 0.40 & 0.50 \end{pmatrix}$	$\begin{pmatrix} 1.00 & 3.00 \\ 3.00 & 4.00 \end{pmatrix}$
2	$\begin{pmatrix} 0.25 & -0.50 \\ -0.40 & -0.60 \end{pmatrix}$	$\begin{pmatrix} 0.30 & -0.70 \\ 0.60 & 0.20 \end{pmatrix}$	$\begin{pmatrix} 9.00 & 4.00 \\ 4.00 & 1.00 \end{pmatrix}$
3	$\begin{pmatrix} -0.20 & -0.70 \\ 0.60 & 0.70 \end{pmatrix}$	$\begin{pmatrix} -0.30 & 0.20 \\ 0.80 & 0.90 \end{pmatrix}$	$\begin{pmatrix} 9.00 & 4.00 \\ 4.00 & 1.00 \end{pmatrix}$
4	$\begin{pmatrix} -0.60 & 0.50 \\ 0.50 & -0.15 \end{pmatrix}$	$\begin{pmatrix} -0.50 & 0.10 \\ 0.70 & 0.90 \end{pmatrix}$	$\begin{pmatrix} 1.00 & 3.00 \\ 3.00 & 4.00 \end{pmatrix}$

,

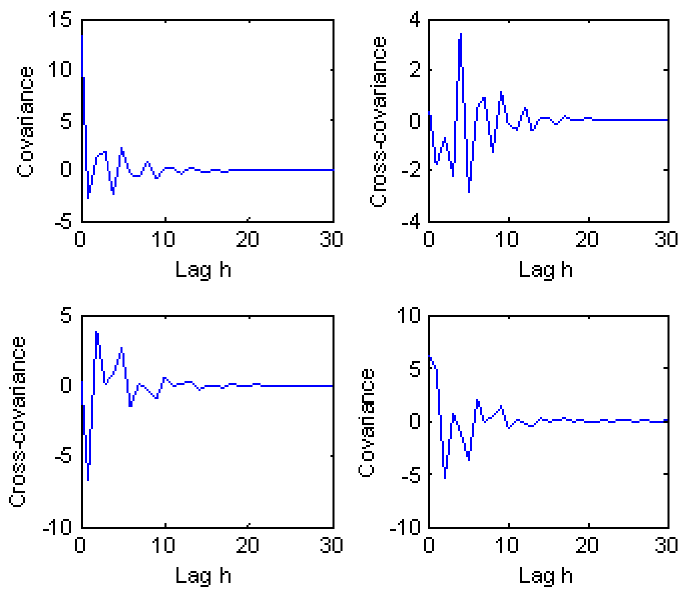
figures $1 - 4$ plot the autocovariances of the above $PVARMA$ model for each of the four seasons.



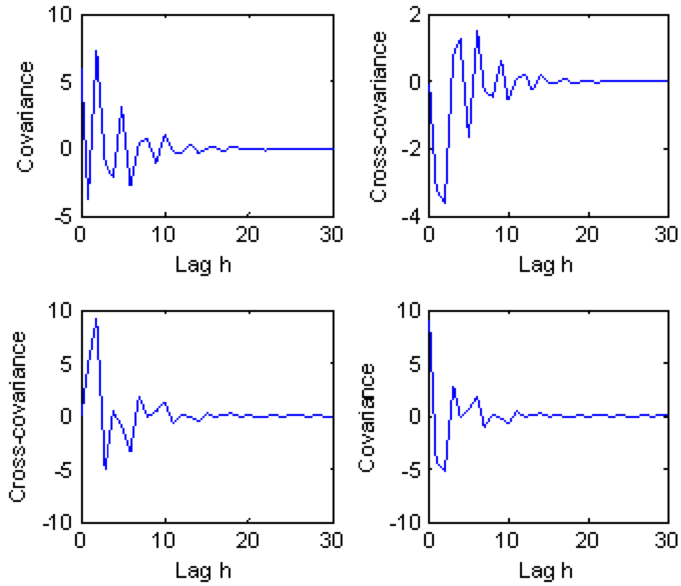
$PVARMA$ Cross-covariances for season 1.



$PVARMA$ Cross-covariances for season 2.



$PVARMA$ Cross-covariances for season 3.



$PVARMA$ Cross-covariances for season 4.

1.3.3 Conclusion

In this section an efficient procedure for obtaining causal periodic $VARMA$ autocovariances has been given. Even for the univariate $PARMA$ case, the proposed method seems to be the most efficient one, especially when the period S is greater than the order p. Adopting Gladyshev's theorem, a higher dimensional periodically stationary $VARMA$ model can be transformed to an equivalent $PVARMA$ model for which the aforementioned procedure can be applied efficiently to obtain its autocovariances. A possible limitation of our algorithm is the need to form explicitly some matrix products and sum of matrix products. This complexity can be avoided by some periodic decomposition procedures, such as the periodic Schur decomposition (e.g., Bojanczyk et al., 1992; Varga 1997).

1.4 On the limiting Fisher information matrix of PARMA models

It is well known that the structure of the Fisher information matrix plays a primordial role in a large variety of estimation methods which are, under the Gaussian assumption, iterative in nature. Indeed, the recognition of its importance and its utilization go back in the statistical analysis history to Whittle (1953), who introduced for the first time the asymptotic matrix information, which is based on the approximation of the Gaussian likelihood. Nowadays, recursive estimation algorithms need for the use of this important information tool (see, Klein et al., 1998; Klein and Mélard, 1989, 1990, 2005; Mélard and Klein, 1994; Porat and Friedlander, 1986*b*; Duflo, 1990; Newton, 1978 and many others).

The quasi-totality of the existing results tied to the calculation of the Fisher information matrices are devoted to the time-invariant coefficients of causal and invertible autoregressive moving average models. The main goal of many works and investigations conducted in this subject is to obtain algorithms which minimize the computation time and keep the needed memory space as small as possible (see, Klein et al., 1998).

In spite of the primary importance and the increasing need for the estimation purpose of periodic time series models, the calculation of Fisher information matrices of these models has received much less interest in the litterature. Indeed, it appears that the only result existing in the literature concerning the calculation of the periodic Fisher information matrix, is the result of Pagano (1978), which is tied to the periodic autoregressive model. Moreover, this result cannot be explored in a straightforward manner because it needs the calculation of the autocovariance function in terms of the periodic model parameters. This section is concerned with the limiting Fisher information matrix of univariate $PARMA$ models.

The rest of this section is organized as follows. We first present an algorithm for the computation of the limiting Fisher information matrix for $PARMA$ models. The established algorithm extends the Klein and Mélard algorithm (1989), elaborated for the classical $ARMA$ models. As the Fisher information matrix is useful by its inverse, we give a necessary and sufficient condition for nonsingular Fisher information matrix for a $PARMA$ model. Finally, we illustrate our method via an example.

Consider the following univariate $PARMA_S(p, q)$ model

$$y_t - \sum_{j=1}^{p} \phi_{tj}\, y_{t-j} = \varepsilon_t - \sum_{j=1}^{q} \theta_{tj}\, \varepsilon_{t-j},\ t \in \mathbb{Z}, \tag{1.4.1}$$

where $\{\varepsilon_t, t \in \mathbb{Z}\}$ is a periodic white noise with variance σ_t^2. Let L denote, as usual, the backward operator, i.e., $L^j\, y_t = y_{t-j},\ j \in \mathbb{N}$, and let $\phi_t(L)$ and $\theta_t(L)$ denote the unidimensional linear difference operators associated, respectively, to the autoregressive and moving average parts of model (1.4.1). Using this notation, one can rewrite (1.4.1) in the condensed form

$$\phi_t(L)\, y_t = \theta_t(L)\, \varepsilon_t,\ t \in \mathbb{Z}. \tag{1.4.2}$$

As in section 1.2, letting $t = i + \tau S$, where $\tau = \left\lceil \frac{t-1}{S} \right\rceil$, $i = 1, ..., S$ and t, $\tau \in \mathbb{Z}$, the periodic model (1.4.1) can then be written as

$$y_{i+\tau S} - \sum_{j=1}^{p} \phi_{i,j}\, y_{i-j+\tau S} = \varepsilon_{i+\tau S} - \sum_{j=1}^{q} \theta_{i,j}\, \varepsilon_{i-j+\tau S}. \tag{1.4.3}$$

Let $\beta' = \left(\underline{\phi}_1', \, \underline{\theta}_1', \, \underline{\phi}_2', \, \underline{\theta}_2', \, ..., \, \underline{\phi}_S', \, \underline{\theta}_S'\right)$ be the $(p+q)\,S \times 1$-vector of the autoregressive and moving average parameters and $\boldsymbol{\sigma}' = (\sigma_1^2, \sigma_2^2, ..., \sigma_S^2)$ be the $S \times 1$-vector of parameter variances, where $\underline{\phi}_i = (\phi_{i,1}, \phi_{i,2} ..., \phi_{i,p})'$ and $\underline{\theta}_i = (\theta_{i,1}, \theta_{i,2} ..., \theta_{i,q})'$, $i = 1, 2, ..., S$. For the parsimony problem, we suppose that the polynomial operators $\phi_t\,(L)$ and $\theta_t\,(L)$ have no common roots. As we can see below, this condition allows the Fisher information matrix of the $PARMA$ model (1.4.3) to be nonsingular. Furthermore, we make the following assumptions :

$A1$. The distribution of the innovation process $\{\varepsilon_t\}$ is supposed Gaussian with periodic variance σ_t^2, i.e., $\varepsilon_t \sim N(0,\, \sigma_t^2)$.

$A2$. The model (1.4.1) is causal and invertible (e.g. the autoregressive and the moving average parameters satisfy, respectively, the necessary and sufficient causality and invertibility conditions (1.2.14) and (1.2.18)).

1.4.1 Fisher's information matrix for $PARMA$ models

Assume that $\underline{y}_{NS}' = (y_1, y_2, ..., y_{NS})$ is a realization of the $PARMA$ model (1.4.3) with a size $M = NS$ supposed without loss of generality multiple of the period S. Recall that the Gaussian Fisher Information Matrix, noted F_{NS}, relative to this time series $\underline{y}_{NS}$ is by definition given by

$$F_{NS}(\beta, \boldsymbol{\sigma}) = -\, E\left[\frac{\partial^2 \log\left(\mathcal{L}\left(\beta, \boldsymbol{\sigma}\right)\right)}{\partial\beta\partial\beta'}\right], \tag{1.4.4}$$

where $\mathcal{L}\left(\beta, \boldsymbol{\sigma}\right)$ is the Gaussian likelihood function associated to the series $\underline{y}_{NS}$. Let $F(\beta, \boldsymbol{\sigma})$ denote the limiting Fisher information matrix (see Whittle, 1953). This last matrix can be partitioned as

$$F(\beta, \boldsymbol{\sigma}) = (F_{jl})_{j,l=1,\ldots,S}$$

where

$$F_{jl} = \begin{pmatrix} F_{\underline{\phi}_j \underline{\phi}_l} & F_{\underline{\phi}_j \underline{\theta}_l} \\ F_{\underline{\theta}_j \underline{\phi}_l} & F_{\underline{\theta}_j \underline{\theta}_l} \end{pmatrix}, \; j, l = 1, ..., S,$$

and the matrices $F_{\underline{\phi}_j \underline{\phi}_l}$, $F_{\underline{\phi}_j \underline{\theta}_l}$, $F_{\underline{\theta}_j \underline{\theta}_l}$ are given by

$$\begin{array}{c} (F_{\underline{\phi}_j\ \underline{\phi}_l})_{km} = F(\phi_{jk}, \phi_{lm}, \boldsymbol{\sigma}) = -\frac{1}{N} E\left[\dfrac{\partial^2 \log \mathcal{L}\left(\beta, \boldsymbol{\sigma}\right)}{\partial \phi_{jk} \partial \phi_{lm}}\right] \\ k = 1, ..., p, m = 1, ..., p \\ (F_{\underline{\phi}_j\ \underline{\theta}_l})_{km} = F(\phi_{jk}, \theta_{lm}, \boldsymbol{\sigma}) = -\frac{1}{N} E\left[\dfrac{\partial^2 \log \mathcal{L}\left(\beta, \boldsymbol{\sigma}\right)}{\partial \phi_{jk} \partial \theta_{lm}}\right] \\ k = 1, ..., p, m = 1, ..., q \\ (F_{\underline{\theta}_j\ \underline{\theta}_l})_{km} = F\left(\theta_{jk}, \theta_{lm}, \boldsymbol{\sigma}\right) = -\frac{1}{N} E\left[\dfrac{\partial^2 \log \mathcal{L}\left(\beta, \boldsymbol{\sigma}\right)}{\partial \theta_{jk} \partial \theta_{lm}}\right] \\ k = 1, ..., q,\ m = 1, ..., q \ \text{ and } j,\ l = 1, ..., S. \end{array} \tag{1.4.5}$$

For the periodic autoregressive case, Pagano (1978) proved that

$$F(\underline{\phi}, \boldsymbol{\sigma}) = \frac{1}{N}\, F_{NS}(\underline{\phi}, \boldsymbol{\sigma}),$$

and showed that the moment estimator $\widetilde{\underline{\phi}}$ of $\underline{\phi}$ is asymptotically normally distributed with zero mean and covariance matrix equal to the inverse of the limiting Fisher information matrix. The expression of this matrix is given by

$$F(\phi_{jk}, \phi_{lm}, \boldsymbol{\sigma}) = \delta_{jl}\, \frac{\gamma_{m-k}^{(i_0)}}{\sigma_k^2}; \quad \begin{array}{l} k,\ m = 1, ..., p \\ j, l = 1, ..., S, \end{array} \tag{1.4.6}$$

where δ_{jl} is the Kronecker function with $j - k = i_0 + rS$, $i_0 = 1, 2, ..., S$ and r is the smallest positive integer such that $1 \leq j - k \leq S$.

For the $PARMA$ case, it seems that there is no corresponding result to (1.4.6). However, in the case of a time-invariant coefficients $ARMA$ model, there exist several adequate algorithms for the computation of the underlying limit information matrix F (see, Klein and Mélard 1989). In this section, we extend this algorithm to the periodic case.

To calculate the elements of the matrix $F(\beta, \boldsymbol{\sigma})$, we associate to the underlying process $\{y_t,$ $t \in \mathbb{Z}\}$ given by (1.4.3) the $S-$periodically correlated adjoint process $\{Z_t,\ t \in \mathbb{Z}\}$, solution to the following PAR difference equation

$$\Phi_i^*\left(L\right)\ Z_{i+S\tau} = \varepsilon_{i+S\tau},\ i = 1, ..., S,\ \tau \in \mathbb{Z} \tag{1.4.7}$$

where

$$\Phi_i^*(L) = \phi_i(L)\,\theta_i\,(L) = 1 - \sum_{j=1}^{P}\phi_{ij}^*\,L^j,$$
$$\phi_{ij}^* = -\sum_{k=0}^{j}\phi_{ik}\,\theta_{i,j-k},\;\; j = 1,...P$$

and $P = p + q$, with the convention $\phi_{ij}, \theta_{ik} = 0$ for $k > q$, $j > p$ or $j, k < 0$.

Furthermore, let $\phi_i^* = (\phi_{i1}^*, \phi_{i2}^*, ..., \phi_{iP}^*)'$, $i = 1, ..., S$, be the $(P \times 1)-$vector of the parameters associated to the i-th period.

This approach, based on Whittle's result (see Whittle, 1953), was suggested and used by McLeod (1984) and Klein and Mélard (1989) for the classical univariate $ARMA$ models and by Pham (1986) for $VARMA$ models. This one can also be used in our periodic case without ambiguity if we suppose that the periodically correlated process Z_t, defined above, admits a spectral factorization (cf., Bentarzi and Hallin, 1998). Using these notations and definitions we can state the following main result which provides the computation of $F(\beta, \boldsymbol{\sigma})$.

Proposition 1.4.1 (Bentarzi and Aknouche, 2005)

The Fisher information matrix of the parameters vector β for a Gaussian $S-$ $PARMA$ model is symmetric block diagonal with elements given by :

$$F(\phi_{jk}, \phi_{lm}, \boldsymbol{\sigma}) = \delta_{jl}\,\sigma_j^{-2}\sum_{s=0}^{q}\sum_{h=0}^{q}\theta_{js}\,\theta_{jh}\,\gamma_{(m+h)-(k+s)}^{(i_0)} \qquad j,\; l = 1, ..., S;\; k \text{ and } m = 1, ..., p; \tag{1.4.8}$$

$$F\left(\theta_{jk}, \theta_{lm}, \boldsymbol{\sigma}\right) = \delta_{jl}\,\sigma_j^{-2}\sum_{s=0}^{p}\sum_{h=0}^{p}\phi_{js}\,\phi_{jh}\,\gamma_{(m+h)-(k+s)}^{(i_0)}, \qquad j,\; l = 1, ..., S;\; k, \text{ and } m = 1, ..., q; \tag{1.4.9}$$

$$F(\theta_{jk}, \phi_{lm}, \boldsymbol{\sigma}) = -\,\delta_{jl}\,\sigma_j^{-2}\sum_{s=0}^{p}\sum_{h=0}^{q}\phi_{js}\,\theta_{jh}\,\gamma_{(m+h)-(k+s)}^{(i_0)}, \qquad j, l = 1, ..., S;\; \text{and } k = 1, ..., q,\; m = 1, ..., p, \tag{1.4.10}$$

where $j - (k + s) = i_0 + rS$, $i_0 = 1, 2, ..., S$ and r is the smallest positive integer such that $1 \leq j - (k + s) \leq S$.

Proof

For the observed time series vector $\underline{y}'_{NS} = (y_1, y_2, ..., y_{NS})$, the logarithm of the conditional likelihood is given by :

$$\log \mathcal{L}(\beta, \boldsymbol{\sigma}) = -\frac{NS \log(2\pi)}{2} - \frac{N}{2}\sum_{i=1}^{S} \log\left(\sigma_i^2\right) - \frac{1}{2}\sum_{\tau=0}^{N-1}\sum_{i=1}^{S}\frac{\varepsilon_{i+S\tau}^2(\beta)}{\sigma_i^2}. \tag{1.4.11}$$

Since $\varepsilon_t(\beta)$ coincides with ε_t under the assumption that $\underline{y}'_{NS}$ is a realization of a process satisfying (1.4.3), we can then recursively calculate $\varepsilon_t(\beta)$, $t = 1, 2, ..., SN$, for the initial values $\varepsilon_t(\beta) = 0$ and $y_t = 0$ for $t \leq 0$. Hence,

$$\frac{\partial^2 \log \mathcal{L}(\beta, \boldsymbol{\sigma})}{\partial\beta\partial\beta'} = -\sum_{i=1}^{S}\frac{1}{\sigma_i^2}\sum_{\tau=0}^{N-1}\left[\frac{\partial\varepsilon_{i+S\tau}(\beta)}{\partial\beta}\frac{\partial\varepsilon_{i+S\tau}(\beta)}{\partial\beta'} + \varepsilon_{i+S\tau}(\beta)\frac{\partial^2\varepsilon_{i+S\tau}(\beta)}{\partial\beta\partial\beta'}\right].$$

We note that the second derivative $\dfrac{\partial^2\varepsilon_t(\beta)}{\partial\beta\partial\beta'}$ can be expressed as a linear combination of only the past values of y_t, then it is uncorrelated with $\varepsilon_t(\beta)$ which coincides with the innovation processes. Consequently, we have the property which we need ulteriorly

$$E\left[\varepsilon_t(\beta)\ \frac{\partial^2\varepsilon_t(\beta)}{\partial\beta\partial\beta'}\right] = 0.$$

From the latter remark we can write the matrix F as follows

$$F(\beta, \boldsymbol{\sigma}) = N^{-1}\sum_{i=1}^{S}\frac{1}{\sigma_i^2}\sum_{\tau=0}^{N-1}\left[\frac{\partial\varepsilon_{i+S\tau}(\beta)}{\partial\beta}\ \frac{\partial\varepsilon_{i+S\tau}(\beta)}{\partial\beta'}\right]$$

and by periodic stationarity

$$F(\beta, \boldsymbol{\sigma}) = \sum_{i=1}^{S}\frac{1}{\sigma_i^2}E\left[\frac{\partial\varepsilon_{i+S\tau}(\beta)}{\partial\beta}\frac{\partial\varepsilon_{i+S\tau}(\beta)}{\partial\beta'}\right] \tag{1.4.12}$$

Hence we have to calculate only the S expressions

$$E\left[\frac{\partial\varepsilon_{i+S\tau}(\beta)}{\partial\beta}\ \frac{\partial\varepsilon_{i+S\tau}(\beta)}{\partial\beta'}\right].$$

The derivatives of $\varepsilon_i(\beta)$ with respect to β can be established from (1.4.2). Indeed,

$$\frac{\partial\varepsilon_{i+S\tau}(\beta)}{\partial\phi_{jk}} = \begin{cases} -L^k\theta_i(L)\ y_{i+S\tau} = -L^k\phi_i^{-1}(L)\ \varepsilon_{i+S\tau}(\beta) & \text{if } j = i \\ 0 & \text{if } j \neq i \end{cases} \tag{1.4.13}$$

and

$$\frac{\partial \varepsilon_{i+S\tau}(\beta)}{\partial \theta_{lm}} = \begin{cases} L^m \theta_i^{-2}(L)\,\phi_i(L)\,y_{i+S\tau} = L^m \theta_i^{-1}(L)\,\varepsilon_{i+S\tau}(\beta), & \text{if } l = i \\ 0, & \text{otherwise} \end{cases} \tag{1.4.14}$$

Consequently, the elements of F can be obtained from (1.4.12) – (1.4.14). More precisely,

$$F\left(\phi_{jk}, \phi_{lm}, \boldsymbol{\sigma}\right) = \sum_{i=1}^{S} \frac{1}{\sigma_i^2} E\left[\frac{\partial \varepsilon_{i+S\tau}(\beta)}{\partial \phi_{jk}} \frac{\partial \varepsilon_{i+S\tau}(\beta)}{\partial \phi_{lm}}\right]$$

$$= \begin{cases} \dfrac{1}{\sigma_j^2} E\left[\dfrac{\partial \varepsilon_{i+S\tau}(\beta)}{\partial \phi_{jk}} \dfrac{\partial \varepsilon_{i+S\tau}(\beta)}{\partial \phi_{lm}}\right] & \text{if } l = j, \\ & j,\ l = 1, ..., S \\ & k,\ m = 1, ..., p. \\ 0 & \text{otherwise} \end{cases}$$

Using (1.4.7) and (1.4.13)

$$F\left(\phi_{jk}, \phi_{lm}, \boldsymbol{\sigma}\right) = \begin{cases} \dfrac{1}{\sigma_j^2}\ E\left(L^k\ \theta_j(L)\ \ Z_j\ L^m\ \theta_j(L)\ \ Z_j\right), \\ \qquad\qquad\qquad\qquad \text{if } l = j \\ 0, \qquad\qquad \text{otherwise} \end{cases}$$

$$= \begin{cases} \delta_{jl} \frac{1}{\sigma_j^2} E\left[\left(\sum_{s=0}^{q} \theta_{js} L^{k+s}\right) Z_j\ \left(\sum_{h=0}^{q} \theta_{jh} L^{m+h}\right) Z_j\right], \\ \qquad\qquad\qquad\qquad \text{if } l = j \\ 0, \qquad \text{otherwise} \end{cases}$$

$$= \delta_{jl}\ \sigma_j^{-2} \sum_{s=0}^{q} \sum_{h=0}^{q} \theta_{js}\ \theta_{jh}\ E\left(Z_{j-(k+s)}\ Z_{j-(m+h)}\right)$$

$$= \delta_{jl}\ \sigma_j^{-2} \sum_{s=0}^{q} \sum_{h=0}^{q} \theta_{js}\ \theta_{jh}\ \gamma_{(m+h)-(k+s)}^{(j-(k+s))},$$

where $\gamma_h^{(i)} = E(Z_{i+S\tau} Z_{i-h+S\tau})$ is the autocovariance function at lag h and season i of the process $\{Z_t,\ t \in \mathbb{Z}\}$. In the same manner,

$$F\left(\theta_{jk}, \theta_{lm}, \boldsymbol{\sigma}\right) = \begin{cases} \frac{1}{\sigma_j^2}\ E\left(L^k \phi_j(L)\, Z_j L^m \phi_j(L)\, Z_j\right) & \text{if } l = j \\ 0, & \text{otherwise} \end{cases}$$

$$= \delta_{jl}\ \frac{1}{\sigma_j^2}\ E\left[\left(\sum_{s=0}^{q}\ \phi_{js}\ L^{k+s}\right) Z_j \left(\sum_{h=0}^{q}\ \phi_{jh}\ L^{m+h}\right) Z_j\right]$$

$$= \delta_{jl}\ \sigma_j^{-2} \sum_{s=0}^{q} \sum_{h=0}^{q} \phi_{js}\ \phi_{jh}\ \gamma_{(m+h)-(k+s)}^{(j-(k+s))}.$$

Finally, from (1.4.7), (1.4.13) and (1.4.14)

$$F\left(\theta_{jk},\phi_{lm},\boldsymbol{\sigma}\right)=\begin{cases}-\dfrac{1}{\sigma_j^2}E(L^k\phi_j\left(L\right)Z_jL^m\theta_j\left(L\right)Z_j, & \text{if } l=j\\ 0, & \text{otherwise}\end{cases}$$

$$=-\,\delta_{jl}\,\sigma_j^{-2}\sum_{s=0}^{q}\sum_{h=0}^{q}\phi_{js}\,\theta_{jh}\,\gamma_{(m+h)-(k+s)}^{(j-(k+s))}.$$

Consequently, we obtain the Fisher information matrix $F(\beta,\boldsymbol{\sigma})$ and its inverse $F^{-1}(\beta,\boldsymbol{\sigma})$, which is block diagonal. ■

In the case where $\underline{\theta}=(\underline{\theta}_1',\underline{\theta}_2',...,\underline{\theta}_q')'=0_{qS\times 1}$ we find again Pagano's result (1978). Note that relations (1.4.8) – (1.4.10) can be given in another form by using the auxiliary matrix J of a $PARMA_S(p,q)$ model, which is defined by (see McLeod (1984) in the $ARMA$ case)

$$J=\underset{i=1,\ldots,S}{diag}\,(J_i),$$

where

$$J_i(j,k)=\begin{cases}\theta_{i,j-k} & \text{if } j=1,...,r,\ k=1,...,p\\ -\phi_{i,j-k} & \text{if } j=1,...,r,\ k=q+1,...,q+p\end{cases}\qquad i=1,...,S,\quad r=\max\left(p,q\right),$$

and $\underset{i=1,\ldots,S}{diag}\,(J_i)$ denotes the block diagonal matrix formed by the matrices J_i. Thus

$$F(\beta,\boldsymbol{\sigma})=JF(\phi^*,\boldsymbol{\sigma})J'.$$

Remark 1.4.1

From (1.4.8) – (1.4.10) we see that the calculation of the $PARMA$ Fisher information matrix involves the computation of the autocovariances of an adjoint PAR model. Thus, estimation of the elements of the $PARMA$ Fisher matrix can be obtained simply by replacing the unknown autocovariances by their corresponding empirical estimations as McLeod (1993) has made (see also Vecchia, 1985), or more efficiently by using the theoretical autocovariances as shown in section 1.3.

In the present subsection, the calculation of the Fisher information matrix for $PARMA$ models has been explored. Next, we study the invertibility property of such a matrix, which is a very useful tool for statistical inferences tied to $PARMA$ models.

1.4.2 Necessary and sufficient condition for nonsingular Fisher information matrix

In this subsection, we give a necessary and sufficient condition for the invertibility of the Fisher information matrix in $PARMA$ models. It is known in the literature of stationary $ARMA$ models that the Fisher information matrix is nonsingular if and only if the corresponding model is non redundant, in the sense that the polynomials AR and MA do have no common roots (see Brockwell and Davis, 1991 Theorem 8.11.1). Poskitt and Treymane (1981) have given a proof of this result, but McLeod (1999) showed it in a much simpler way. The following theorem is inspired from this one.

Theorem 1.4.1 (Bentarzi and Aknouche, 2005)

The Fisher information matrix $F(\beta, \boldsymbol{\sigma})$ of the $PARMA$ model (1.4.3) *is nonsingular if and only if the model* (1.4.3) *is non redundant.*

The proof of this theorem is based on the following lemma.

Lemma 1.4.1 (Bentarzi and Aknouche, 2005)

A necessary and sufficient condition for redundancy of model (1.4.3) *is that there exist* $2.S$ *polynomials* $\alpha_i(L) = \prod_{j=1}^{p-1}(1 - E_{ij}L)$ *and* $\beta_i(L) = \prod_{j=1}^{q-1}(1 - F_{ij}L)$, $i = 1, ..., S$, *such that*

$$\sum_{i=1}^{S} \frac{\alpha_i(L)}{\phi_i(L)} = \sum_{i=1}^{S} \frac{\beta_i(L)}{\theta_i(L)} \tag{1.4.15}$$

Proof

For $i = 1, ..., S$, let $\phi_i(L) = \prod_{j=1}^{p}(1 - G_{ij}L)$ and $\theta_i(L) = \prod_{j=1}^{q}(1 - H_{ij}L)$, then (1.4.15) becomes

$$\sum_{i=1}^{S} \left(\frac{\prod_{k=1}^{p-1}(1 - E_{ik}L)}{\prod_{j=1}^{p}(1 - G_{ij}L)} \right) = \sum_{i=1}^{S} \left(\frac{\prod_{k=1}^{q-1}(1 - F_{ik}L)}{\prod_{j=1}^{q}(1 - H_{ij}L)} \right). \tag{1.4.16}$$

If $H_{i1} = G_{i1}$ for all i, $i = 1, ..., S$, then (1.4.16) holds immediately by setting $E_{ik} = G_{i,k+1}$, $k = 1, ..., p-1$ and $F_{ik} = H_{i,k+1}$, $k = 1, ..., p-1$, $i = 1, ..., S$. Conversely, equation (1.4.16) can be written as follows

$$\left(\frac{\prod_{i=1}^{S} \prod_{j=1}^{q} (1 - H_{ij} L)}{\prod_{i=1}^{S} \prod_{j=1}^{p} (1 - G_{ij} L)} \right) = \frac{B(L)}{A(L)} \tag{1.4.17}$$

where $A(L)$ and $B(L)$ are two polynomials of degrees $(p-1)S$ and $(q-1)S$ respectively. As $H_{ij} \neq G_{ik}$ for all i, j, k, the left member of (1.4.17) is then irreducible, whose numerator and denominator are of degrees pS and qS, respectively. Consequently, identity (1.4.17) can never hold taking into account the degrees $(p-1)S$ and $(q-1)S$ of the numerator and denominator in the right member of (1.4.17). ■

Proof of Theorem 1.4.1

The matrix F in (1.4.12) can be written as

$$F(\beta, \boldsymbol{\sigma}) = \sum_{i=1}^{S} \frac{1}{\sigma_i^2} E\left[A_{i+S\tau} A'_{i+S\tau} \right], \tag{1.4.18}$$

where

$$A_{i+S\tau} = \begin{cases} (\underline{v}'_{1,\tau}, \underline{u}_{1,\tau}, \underline{0}', ..., \underline{0}')', & if \;\; i = 1, \\ (\underline{0}', ..., \underline{0}', \underline{v}'_{i,\tau}, \underline{u}_{i,\tau}, \underline{0}', ..., \underline{0}')', & if \;\; 2 \leq i \leq S-1, \\ (\underline{0}', \underline{0}' ..., \underline{0}', \underline{v}'_{S,\tau}, \underline{u}_{S,\tau})', & if \;\; i = S, \end{cases} \tag{1.4.19}$$

and where the $S \times (p+q)$ vectors $\underline{v}_{i,\tau} = (v_{i+S\tau-1}, v_{i+S\tau-2}, ..., v_{i+S\tau-p})'$ and $\underline{u}_{i,\tau} = (u_{i+S\tau-1}, u_{i+S\tau-2}, ..., u_{i+S\tau-p})'$ are, respectively, the vector of derivative of $\varepsilon_{i+S\tau}(\beta)$ with respect to $\underline{\phi}$ and to $\underline{\theta}$, with

$$\begin{aligned} \phi_i(L)\, v_{i+S\tau} &= \varepsilon_{i+S\tau}, \\ \theta_i(L)\, u_{i+S\tau} &= -\varepsilon_{i+S\tau}, \end{aligned} \tag{1.4.20}$$

(In the $ARMA$ case, this formula was given by Box and Jenkins (1976), p.241). However, $F(\beta, \boldsymbol{\sigma})$ is singular if and only if there exist $x \in R^{(p+q)S}$, such that : $x' F(\beta, \boldsymbol{\sigma}) x = 0$. In other words, $\exists \alpha_{ij}, \beta_{ik}$, $i = 1, .., S$, $j = 0, ..., p-1$ and $k = 0, ..., q-1$, such that

$$\sum_{i=1}^{S} \frac{1}{\sigma_i^2} Var \left(\sum_{j=0}^{p-1} \alpha_{ij} v_{i+S\tau-j} + \sum_{k=0}^{q-1} \beta_{ik} u_{i+S\tau-k} \right) = 0. \tag{1.4.21}$$

Taking account of (1.4.13), (1.4.14) and (1.4.20), equation (1.4.21) is equivalent to

$$Var\sum_{i=1}^{S}\frac{1}{\sigma_i}\left(\sum_{j=0}^{p-1}\alpha_{ij}v_{i+S\tau-j}+\sum_{k=0}^{q-1}\beta_{ik}u_{i+S\tau-k}\right)=0,$$

or to the condition that there exist $2S$ polynomials $\alpha_i(L)=\sum_{j=0}^{p-1}\alpha_{ij}L^i$ and $\beta_i(L)=\sum_{j=0}^{p-1}\beta_{ij}L^i$, $i=1,..,S$ such that

$$\sum_{i=1}^{S}\frac{1}{\sigma_i}\left(\alpha_i(L)v_{i+S\tau-1}+\beta_i(L)u_{i+S\tau-1}\right)=0. \tag{1.4.22}$$

By using (1.4.20), equation (1.4.22) can be written as

$$\sum_{i=1}^{S}\frac{1}{\sigma_i}\left(\frac{\alpha_i(L)}{\phi_i(L)}\varepsilon_{i+S\tau}-\frac{\beta_i(L)}{\theta_i(L)}\varepsilon_{i+S\tau}\right)=0. \tag{1.4.23}$$

For $i=1,...,S$, multiplying equation (1.4.23) by $\varepsilon_{i+S\tau}$, taking expectation and then summing the S obtained equations, we get

$$\sum_{i=1}^{S}\frac{\alpha_i(L)}{\phi_i(L)}=\sum_{i=1}^{S}\frac{\beta_i(L)}{\theta_i(L)},$$

which by the previous lemma is a necessary and sufficient condition for the redundancy of model (1.4.3). ■

1.4.3 Illustrative example

The following example is devoted to a simple illustration of the application of the computation procedures presented in the previous sections. We consider a 4−periodic autoregressive moving average of order $(3,2)$ with $\underline{\phi}_1'$, $\underline{\phi}_2'$, $\underline{\phi}_3'$, $\underline{\phi}_4'$ are respectively $(.1,-.5,.2)'$, $(.5,.4,.8)'$, $(.1,.2,.3)'$, $(-.4,-.2,.9)'$; $\underline{\theta}_1'$, $\underline{\theta}_2'$, $\underline{\theta}_3'$, $\underline{\theta}_4'$ being $(.2,.3)'$, $(.1,.3)'$, $(.1,.8)'$, $(-.5,.5)'$ and $\boldsymbol{\sigma}=(1,2,1,1)'$. The Fisher information matrix is then given by $F=\underset{i=1,...,S}{diag}(Fi)$ where

$$F_1=\begin{pmatrix} 3.1530 & -1.0470 & -1.4221 & .3782 & 1.3282\\ & 2.3826 & .2078 & 2.3634 & 1.6994\\ & & 3.1444 & .6240 & 2.8895\\ & & & 1.5738 & .6996\\ & & & & 3.3971\end{pmatrix},$$

$$F_2 = \begin{pmatrix} .9270 & .3328 & -.6286 & -1.3504 & -.7445 \\ & 1.4605 & -.4427 & -1.3701 & -.8031 \\ & & 1.2861 & -.8677 & .2876 \\ & & & 1.4455 & .8163 \\ & & & & 2.4958 \end{pmatrix},$$

$$F_3 = \begin{pmatrix} 2.8074 & -.8031 & -.6444 & -.0350 & -.4635 \\ & 4.1373 & 1.5205 & -1.0410 & -.3262 \\ & & 5.2450 & -1.0100 & -1.2100 \\ & & & 2.2652 & .8427 \\ & & & & 3.0309 \end{pmatrix},$$

$$F_4 = \begin{pmatrix} 4.8660 & 3.2234 & 2.8248 & .1180 & -.6237 \\ & 4.1320 & 1.3849 & 2.0551 & -2.2753 \\ & & 4.1781 & 2.3062 & -2.4219 \\ & & & 2.1559 & -.4021 \\ & & & & 4.2843 \end{pmatrix}$$

1.4.4 Conclusion

The aforementioned algorithm makes it possible to obtain the elements of $PARMA$ Gaussian Fisher information matrix for only the autoregressive and moving average parameters. The presented algorithm extends both Pagano's result concerning the the Fisher information matrix for PAR models and the Klein and Mélard (1989) relative to stationary $ARMA$ models. We saw that this matrix is based on the expression of the Gaussian conditional likelihood. This algorithm is recommended for its simplicity and for the fact that for a long time series, this one will be very close to the exact information matrix. We finally note that the adaptation of results (1.4.8)-(1.4.10) to the multivariate $PVARMA$ case can be made straightforwardly.

1.5 Summary

In this chapter, $PVARMA$ models and some of their algebraic and probabilistic properties have been studied. Our contribution is threefold. First, the causality and invertibility problems for $PVARMA$ models have been considered. We have proposed some necessary and sufficient causality and invertibility conditions that seem computationally and analytically simple to check. Second, we have considered the autocorrelation structure of $PVARMA$ models. We developed a computational procedure for $PVARMA$ autocovariances, that compute the autocovariances for distinct seasons separately. It has been shown that the autocovariance calculation reduces to solve a discrete time periodic Lyapunov equation ($DTPLE$). The proposed method can be improved by taking into account the particular periodic structure of the variable of the $DTPLE$. Finally, we have studied the limiting Fisher information matrix for $PARMA$ models. An algorithm for computing the limiting Fisher information matrix for $PARMA$ models has been developed. The proposed algorithm uses the $PARMA$ autocovariance calculation procedure as a subroutine. We have also given a necessary and sufficient condition for the invertibility of such a matrix in terms of the model parameters. As a possible future research, it is interesting to extend the algorithm to the multivariate case. Moreover, the calculation of the exact Fisher information matrix seems not to have been explored until there.

Chapter 2

Off-line estimation of periodic linear models

2.1 Introduction

Estimation methods for time series models can be classified into two general categories, according to the type of data on which they are based : off-line methods and on-line methods. In the first case, time series data are of fixed size, whereas in the second case, they are gradually available in real time. The present chapter is concerned with off-line estimation methods. on-line estimation algorithms for periodic time series models will be treated in the third and the fifth chapters.

Off-line estimation techniques for $PARMA$ models have been widely studied in the last two decades. Existing investigations have been confined to a variety of methods and approaches, such as the method of moment (Pagano, 1978; Boshnakov; 1996; Anderson and Meerschaert, 2005), the Bayesian approach (Andel, 1983), the least squares technique (Adams and Goodwin, 1995; Basawa and Lund, 2001; Bentarzi and Aknouche, 2006) and the maximum likelihood method (Vecchia, 1985a; Li and Hui, 1988; Jymenez et al., 1989; Lund and Basawa, 2000, Basawa and Lund, 2001). These techniques can also be classified into two general classes : methods that use the data directly; and methods that apply some preliminary transformations to the data (see Porat and Friedlander (1986a) for the $ARMA$ case). Among the methods in the first class, we mention in particular the exact maximum likelihood and its many approximations (conditional, unconditional) and the least squares method with its particular variants (ordinary, weighted, conditional, unconditional). Typically, these methods are aimed at preserving some desirable asymptotic properties, namely consistency, asymptotic efficiency and asymptotic normality, while reducing their computational complexity. Among the second class of methods, the most common approach, known as the method of moments, is to transform the data into a finite set of statistics and then estimate $PARMA$ parameters from these statistics. As a special case, we mention the periodic Yule-Walker method (Pagano 1978) and its order-recursive Durbin-Levinson formulations (Sakai, 1982; Boshnakov, 1996). Another well-known moments technique is the innovations algorithm (Anderson et al., 1999; Anderson and Meerschaert, 2005). In the special case of PAR models, methods of the second class (methods of moment) are known to be asymptotically equivalent to the maximum likelihood and the weighted least squares methods (Pagano, 1978; Basawa and Lund, 2001; Bentarzi and Aknouche, 2006) which generally provide asymptotically efficient estimates. However, with the presence of a moving average component, $ARMA$ parameter estimation methods based on sample covariances are known to be less efficient than the direct methods such as the maximum likelihood method (cf., Porat and Friedlander, 1986a). This result can be easily generalized to the $PARMA$ case.

As we can see, most of the investigations relating to periodic linear models have been confined to the univariate case. In this chapter we are mainly concerned with off-line estimation methods for periodic $VARMA$ models. Our study will be restricted to the class of methods that use the data directly, such as the maximum likelihood and the least squares methods.

The following section deals with the likelihood evaluation of $PVARMA$ models. We exploit the autocovariance calculation method for $PVARMA$ given in Section 1.3 to evaluate the $PVARMA$ likelihood by means of the Kalman filter. An alternative technique based on the Ljung and Box (1979) method for the particular class $\Im_{11}$ introduced in Bentarzi (1995) (see Section 1.2) is also proposed. Finally, in Section 3 we study the large sample properties of parameter estimates for $PVARMA$ models.

2.2 Maximum likelihood estimation of $PVARMA$ models

Under some suitable conditions, the maximum likelihood method is known to be statistically the most efficient method. However, because of its iterative nature, which requires several likelihood evaluations, it is computationally cumbersome, particularly for Periodic Vector $ARMA$ models ($PVARMA$), which generally retain a considerable number of parameters. Although the likelihood evaluation problem of univariate $PARMA$ models has been considered in the literature (see the previous Section), the general multivariate $PVARMA$ case remains unexplored. The main goal of this section is to develop a fast algorithm for $PVARMA$ likelihood evaluation by means of the

Kalman filter. Our algorithm differs from that of Jymenez et al. (1989) by the starting Kalman recursions which is the key of our algorithm efficiency. This algorithm is presented in the first subsection. For the particular class $\Im_{11}$ of univariate $PARMA$ models, we exploit the particular form of the matrix $\underline{\Phi}$ to derive an efficient algorithm for the exact likelihood evaluation.

2.2.1 Likelihood evaluation of $PVARMA$ models

It is well known that the Kalman filtering technique is the most efficient tool, on the basis of the theoretical operation counts criterion, for likelihood evaluation of time-invariant $ARMA$ models (e.g. Harvey and Phillips, 1978; Gardner et al., 1979; Pearlman, 1980; Mélard, 1984; Kohn and Ansley, 1985). This technique has also been exploited by Shea (1989) for evaluating the likelihood function of $VARMA$ models and by Jymenez et al. (1989) for univariate periodic $ARMA$ models. This subsection develops a fast algorithm for the exact likelihood evaluation of Gaussian $PVARMA$ models, by means of the Kalman filter. We shall consider a special equivalent state-space representation of the underlying model, from which the likelihood function is efficiently derived using the Kalman recursion and the $PVARMA$ autocovariances given in Section 1.3.

Consider the following time-invariant orders $PVARMA$ model

$$\mathbf{y}_t - \sum_{j=1}^{p} \boldsymbol{\phi}_{t,j}\mathbf{y}_{t-j} = \boldsymbol{\varepsilon}_t - \sum_{j=1}^{q} \boldsymbol{\theta}_{t,j}\boldsymbol{\varepsilon}_{t-j} \tag{2.2.1}$$

where $\{\boldsymbol{\varepsilon}_t, t \in \mathbb{Z}\}$ is an $m-$variate periodic white noise process with covariance matrix $E\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right) = \Sigma_t$. Note that the orders p and q can also be considered periodic in time (p_t and q_t) in order to reduce the number of parameters to be estimated. However, we need in this case a constrained estimation procedure, which is not the objective of this chapter. Furthermore, assume that $A1$ and $A2$ given in Section 1.4 hold.

For a given realization $\mathbf{y} = (\mathbf{y}_1', \mathbf{y}_2', ..., \mathbf{y}_{NS}')'$ of model (2.2.1), the likelihood of the parameter vector

$$\boldsymbol{\beta} = \left(vec\left(\boldsymbol{\phi}\right)', vec\left(\boldsymbol{\theta}\right)', vec\left(\boldsymbol{\Sigma}\right)'\right)'$$

can be expressed as follows

$$\mathcal{L}(\boldsymbol{\beta}; \mathbf{y}) = (2\pi)^{-NS/2} \left|\det(\boldsymbol{\Gamma})\right|^{-1/2} \exp\left\{-\frac{1}{2}\mathbf{y}'\boldsymbol{\Gamma}^{-1}\mathbf{y}\right\}, \tag{2.2.2}$$

where

$$\boldsymbol{\phi} = \begin{pmatrix} \boldsymbol{\phi}_{1,1} & \boldsymbol{\phi}_{1,2} & \cdots & \boldsymbol{\phi}_{1,p} \\ \boldsymbol{\phi}_{2,1} & \boldsymbol{\phi}_{2,2} & \cdots & \boldsymbol{\phi}_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\phi}_{S,1} & \boldsymbol{\phi}_{S,2} & \cdots & \boldsymbol{\phi}_{S,p} \end{pmatrix}, \boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\theta}_{1,1} & \boldsymbol{\theta}_{1,2} & \cdots & \boldsymbol{\theta}_{1,q} \\ \boldsymbol{\theta}_{2,1} & \boldsymbol{\theta}_{2,2} & \cdots & \boldsymbol{\theta}_{2,q} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\theta}_{S,1} & \boldsymbol{\theta}_{S,2} & \cdots & \boldsymbol{\theta}_{S,q} \end{pmatrix},$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_2 & \cdots & \boldsymbol{\Sigma}_S \end{pmatrix},$$

and Γ is the covariance matrix of $\mathbf{y}$.

Let $\widehat{\mathbf{y}}_{s+nS/s+nS-1}$ be the best linear predictor of $\mathbf{y}_{s+nS}$ based on $\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_{s+nS-1}$, (i.e. the vector formed by the orthogonal projection of the entries of $\mathbf{y}_{s+nS}$ in the subspace spanned by $\{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_{s+nS-1}\}$) and $\widehat{\boldsymbol{\varepsilon}}_{s+nS} = \mathbf{y}_{s+nS} - \widehat{\mathbf{y}}_{s+nS/s+nS-1}$ be the sample innovation at time $s+nS$ with mean square error

$$\mathbf{h}_{s+nS} = E\left(\mathbf{y}_{s+nS} - \widehat{\mathbf{y}}_{s+nS/s+nS-1}\right)\left(\mathbf{y}_{s+nS} - \widehat{\mathbf{y}}_{s+nS/s+nS-1}\right)'.$$

Because the normality assumption, $\{\widehat{\boldsymbol{\varepsilon}}_{s+nS}\}$ is a sequence of independent Gaussian random vectors. Furthermore, the innovations vector $\widehat{\boldsymbol{\varepsilon}} = (\widehat{\boldsymbol{\varepsilon}}_1', \widehat{\boldsymbol{\varepsilon}}_2', ..., \widehat{\boldsymbol{\varepsilon}}_{NS}')'$ and the observations vector $\mathbf{y}$ are in a one-to-one correspondence with unit Jacobian. Therefore, the likelihood given by (2.2.2) can be written (cf, Schweppe, 1965) in the innovation form

$$\mathcal{L}(\boldsymbol{\beta}; \mathbf{y}) = (2\pi)^{-\frac{NS}{2}} \prod_{s=1}^{S} \prod_{n=0}^{N-1} (\det \mathbf{h}_{s+nS})^{-1/2} \exp\left\{ -\frac{1}{2} \sum_{s=1}^{S} \sum_{n=0}^{N-1} \widehat{\boldsymbol{\varepsilon}}_{s+nS}' \mathbf{h}_{s+nS}^{-1} \widehat{\boldsymbol{\varepsilon}}_{s+nS} \right\}, \tag{2.2.3}$$

in which we need to evaluate $\mathbf{h}_{s+nS}$, $\widehat{\mathbf{y}}_{s+nS/s+nS-1}$ and hence $\widehat{\boldsymbol{\varepsilon}}_{s+nS}$ for $s = 1, ..., S$ and $n = 0, ..., N-1$. This can be achieved recursively by using the well-known Kalman Filter, provided that we can express model (2.2.1) in a state space form.

Among several state space forms that can be used (e.g. Jymenez et al., 1989; Mélard, 1985; Azrak and Mélard, 1998) to represent a $PVARMA$ model, we choose the well-known $\max(p, q+1)$ representation (e.g. De Jong and Penzer, 2004), which involves fewer operations. Let $r = \max(p, q+1)$ and define the rm-variate process $\{\mathbf{X}_t,\ t \in\}$ as follows

$$\mathbf{X}_{s+nS}(i) = \sum_{j=i}^{r} \boldsymbol{\phi}_{s+nS+i-1,j} \mathbf{y}_{s+nS+i-j-1} - \boldsymbol{\theta}_{s+nS+i-1,j-1} \boldsymbol{\varepsilon}_{s+nS+i-j}, \qquad i = 1, .., r,\ \ s = 1, ..., S \text{ and } n = 0, ..., N-1. \tag{2.2.4}$$

Then model (2.2.1) can be represented in the following state space form (see, e.g. Jymenez et al., 1989 and Azrak and Mélard, 1998)

$$\begin{cases} \mathbf{X}_{s+nS} = \mathbf{F}_{s+nS}\mathbf{X}_{s+nS-1} + \mathbf{G}_{s+nS}\boldsymbol{\varepsilon}_{s+nS} \\ \mathbf{y}_{s+nS} = \mathbf{H}\mathbf{X}_{s+nS} \end{cases}, \tag{2.2.5}$$

for some initial state $\mathbf{X}_0$ with zero mean and covariance matrix $\mathbf{W}_0$, where

$$\mathbf{F}_s = \begin{pmatrix} \boldsymbol{\phi}_{s,1} & \mathbf{I}_{m\times m} & \cdots & \mathbf{0}_{m\times m} \\ \boldsymbol{\phi}_{s+1,2} & \mathbf{0}_{m\times m} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \mathbf{I}_{m\times m} \\ \boldsymbol{\phi}_{s+r-1,r} & \mathbf{0}_{m\times m} & \cdots & \mathbf{0}_{m\times m} \end{pmatrix}, \ \mathbf{G}_s = \begin{pmatrix} \mathbf{I}_{m\times m} \\ -\boldsymbol{\theta}_{s+1,1} \\ \vdots \\ -\boldsymbol{\theta}_{s+r-1,r-1} \end{pmatrix},$$

and

$$\mathbf{H} = \begin{bmatrix} \mathbf{I}_{m\times m} & \mathbf{0}_{m\times m} & \cdots & \mathbf{0}_{m\times m} \end{bmatrix}.$$

From (2.2.5), the orthogonal projection $\widehat{\mathbf{y}}_{s+nS/s+nS-1}$ can be obtained as follows

$$\widehat{\mathbf{y}}_{s+nS/s+nS-1} = \mathbf{H}\widehat{\mathbf{X}}_{s+nS/s+nS-1},$$

where $\widehat{\mathbf{X}}_{s+nS/s+nS-1}$ is the best linear predictor of $\mathbf{X}_{s+nS}$ based on $\mathbf{y}_1, ..., \mathbf{y}_{s+nS-1}$, with mean square error covariance matrix

$$\mathbf{P}_{s+nS/s+nS-1} = E\left(\mathbf{X}_{s+nS} - \widehat{\mathbf{X}}_{s+nS/s+nS-1}\right)\left(\mathbf{X}_{s+nS} - \widehat{\mathbf{X}}_{s+nS/s+nS-1}\right)'.$$

Returning now to the calculation of the $PVARMA$ likelihood given by (2.2.3). As is well known, it is possible to evaluate recursively $\mathbf{h}_{s+nS}$ and $\widehat{\boldsymbol{\varepsilon}}_{s+nS}$ involved in (2.2.3) by means of the Kalman recursion (cf, Kalman, 1960), which gives

$$\begin{aligned} &\widehat{\boldsymbol{\varepsilon}}_{s+nS} = \mathbf{y}_{s+nS} - \mathbf{H}\widehat{\mathbf{X}}_{s+nS/s+nS-1}, \\ &\mathbf{h}_{s+nS} = \mathbf{H}\ \mathbf{P}_{s+nS/s+nS-1}\mathbf{H}', \\ &\mathbf{K}_{s+nS} = \mathbf{F}_{s+1}\mathbf{P}_{s+nS/s+nS-1}\mathbf{H}'\mathbf{h}^{-1}_{s+nS}, \\ &\widehat{\mathbf{X}}_{s+nS+1/s+nS} = \mathbf{F}_{s+1}\widehat{\mathbf{X}}_{s+nS/s+nS-1} + \mathbf{K}_{s+nS}\widehat{\boldsymbol{\varepsilon}}_{s+nS}, \\ &\mathbf{P}_{s+nS+1/s+nS} = \mathbf{F}_{s+1}\mathbf{P}_{s+nS/s+nS-1}\mathbf{F}'_{s+1} + \mathbf{G}_{s+1}\boldsymbol{\Sigma}_{s+1}\mathbf{G}'_{s+1} - \mathbf{K}_{s+nS}\mathbf{h}_{s+nS}\mathbf{K}'_{s+nS}, \end{aligned} \tag{2.2.6}$$

with start-up values

$$\widehat{\mathbf{X}}_{1/0} = \mathbf{0}_{rm\times 1}$$

and

$$\mathbf{P}_{1/0} = \mathbf{F}_1\mathbf{W}_0\mathbf{F}'_1 + \mathbf{G}_1\boldsymbol{\Sigma}_1\mathbf{G}'_1, \tag{2.2.7}$$

where $\mathbf{\Sigma}_s = E\left(\boldsymbol{\varepsilon}_{s+nS}\boldsymbol{\varepsilon}'_{s+nS}\right)$ and $\mathbf{W}_0 = E\left(\mathbf{X}_0\mathbf{X}'_0\right)$. The latter covariance matrix can be efficiently derived from (2.2.4) as follows

$$\begin{aligned}(\mathbf{W}_0)_{i,j} = \sum_{k=i}^{r}\sum_{h=j}^{r}\boldsymbol{\phi}_{i-1,k}\mathbf{\Gamma}^{(i-k-1)}_{i-k-j+h}\boldsymbol{\phi}'_{j-1,h} - \sum_{k=i}^{r}\sum_{h=j}^{r}\boldsymbol{\phi}_{i-1,k}\mathbf{\Psi}_{i-k-1,i-j+h-k-1}\mathbf{\Sigma}_{j-h}\boldsymbol{\theta}'_{j-1,h-1} \\ - \sum_{k=i}^{r}\sum_{h=j}^{r}\boldsymbol{\theta}_{i-1,k-1}\mathbf{\Sigma}_{i-k}\mathbf{\Psi}'_{j-h-1,j-h-i+k-1}\boldsymbol{\phi}'_{j-1,h} + \sum_{k=i}^{r}\boldsymbol{\theta}_{i-1,k-1}\mathbf{\Sigma}_{i-k}\boldsymbol{\theta}'_{j-1,j+k-i-1}.\end{aligned} \tag{2.2.8}$$

Note that the calculation of the autocovariances $\mathbf{\Gamma}^{(s)}_h$ required in (2.2.8) can be easily obtained from the method given in Section 1.3.

Remark 2.2.1

Instead of using (2.2.7) and (2.2.8), the starting covariance matrix $\mathbf{P}_{1/0}$ can be derived by solving a $DTPLE$ of the form

$$\mathbf{P}_{1/0} = \mathbf{F}_1\mathbf{P}_{0/-1}\mathbf{F}'_1 + \mathbf{G}_1\mathbf{\Sigma}_1\mathbf{G}'_1,$$

on $\mathbf{P}_{1/0}$ as done by Jymenez et al. (1989, p. 235). However, the former equation is of order $m\max(p+1,q)$ and it is more suitable to use the (2.2.8) which is based on $\mathbf{\Gamma}^{(S)}_0$, solution of a $DTPLE$ of order mp (cf, Section 1.3). A similar approach for the $ARMA$ and the time-varying $ARMA$ cases has been used, respectively, by Mélard (1984) and Azrak and Mélard (1998).

The following algorithm summarizes the maximum likelihood procedure for $PVARMA$ models.

Algorithm 2.2.1 (Aknouche and Hamdi, 2005)

i) Start with initial estimates $\widehat{\boldsymbol{\beta}}^{(0)} = \left(vec\left(\widehat{\boldsymbol{\phi}}^{(0)}\right)', vec\left(\widehat{\boldsymbol{\theta}}^{(0)}\right)', vec\left(\widehat{\mathbf{\Sigma}}^{(0)}\right)'\right)'$ of

$$\boldsymbol{\beta} = \left(vec\left(\boldsymbol{\phi}\right)', vec\left(\boldsymbol{\theta}\right)', vec\left(\mathbf{\Sigma}\right)'\right)'.$$

ii) Compute $\mathbf{W}_0 = E\left(\mathbf{X}_0\mathbf{X}'_0\right)$ from (2.2.7), where $\mathbf{\Gamma}^{(s)}_h$ is calculated from Algorithm 1.3.1.

iii) Compute the starting value $\mathbf{P}_{1/0} = \mathbf{F}_1\mathbf{W}_0\mathbf{F}'_1 + \mathbf{G}_1\mathbf{\Sigma}_1\mathbf{G}'_1$ of the Kalman recursion (2.2.6).

iv) Compute $\mathbf{h}_{s+nS}$ and $\widehat{\boldsymbol{\varepsilon}}_{s+nS}$ for $s = 1, ..., S$ and $n = 0, ..., N-1$ via the Kalman recursion (2.2.6).

v) Use an optimization routine to find $\widehat{\boldsymbol{\beta}}$ that minimize $-2\mathcal{L}(\boldsymbol{\beta};\mathbf{y})$. The likelihood $\mathcal{L}(\boldsymbol{\beta};\mathbf{y})$ can be evaluated from (2.2.3). ■

Note that the proposed likelihood evaluation algorithm for $PVARMA$ models is no more complex than the evaluation of the stationary $VARMA$ likelihood (cf, Shea, 1989). However, some useful devices such as the Chandrasekhar type recursion (Morf et al., 1974) are needed to improve Kalman recursions for periodic linear systems.

2.2.2 Likelihood evaluation of the particular class $\Im_{11}$

Under the normality assumption, the part which follows deals with the exact likelihood evaluation of the univariate class, $\Im_{11}$, studied in Section 1.2. One thus considers the S-variate $ARMA$ model of order $(1,1)$ equivalent to the $PARMA$ of departure. Although the Kalman filtering technique for likelihood evaluation is considered as the most computational efficient method on the basis of the theoretical operation counts criterion, as mentioned in Mauricio (2002), this technique requires a significant number of preliminary operations to start the filtering recursion, which is not counted in the total number of operations. With regard to the number of preliminary operations, Mauricio (2002) adapted for the $VARMA$ case, the so-called innovations method of Ansley (1979) which is computationally equivalent to the Ljung and Box (1979) method. Mauricio (2002) has shown that the obtained method is the most efficient one and performs better than the Shea (1989) Kalman filter algorithm. Under these last considerations, while basing oneself primarily on the direct method of Ljung and Box (1979) which requires to estimate initial values, one will then exploit the particular form of the matrix $\underline{\Phi}$ (cf, Section 1.2) and an autocovariances calculation method given below, to derive an efficient algorithm for evaluating the exact likelihood. Consider the $PARMA_S(p_t, q_t)$ model

$$y_t - \sum_{j=1}^{p_t} \phi_{t,j} y_{t-j} = \varepsilon_t - \sum_{j=1}^{q_t} \theta_{t,j} \varepsilon_{t-j}, \tag{2.2.9a}$$

belonging to the class $\Im_{11}$ with corresponding $VARMA(1,1)$ of the form

$$\underline{X}_\tau - \underline{\Phi X}_{\tau-1} = \underline{\Theta}_0 \underline{\epsilon}_\tau - \underline{\Theta}_1 \underline{\epsilon}_{\tau-1}, \tag{2.2.9b}$$

where $E\left(\underline{\epsilon}_\tau \underline{\epsilon}'_\tau\right) = diag(\sigma_1^2, \sigma_2^2, ..., \sigma_S^2) = \sigma^2 D$ with $\sigma^2 > 0$ and D is an $S-$diagonal matrix. The specification $\sigma^2 D$ as the covariance matrix of $\underline{\epsilon}_\tau$ is just a useful device for the evaluation of the concentrated likelihood function as we will show below.

Let $\underline{y}'_{NS} = (y_1, y_2, ..., y_{NS})$ be a particular realization of model (2.2.9b). If we put

$$\underline{X}_i = \left(y_{1+(i-1)S}, y_{2+(i-1)S}, ...y_{S+(i-1)S}\right)',$$

and

$$\underline{\epsilon}_i = (\varepsilon_{1+(i-1)S}, \varepsilon_{2+(i-1)S}, ..., \varepsilon_{S+(i-1)S})',$$

for $i = 1, ..., N$, then the series $\underline{y}'_{NS}$ can be written as : $\underline{X}_1, \underline{X}_2, ..., \underline{X}_N$. By rewriting equation (2.2.9b), for $i = 1, ...N$, we obtain the following matrix relation

$$A\underline{X} = B^{-1}\underline{\epsilon} + C\underline{v}, \tag{2.2.10}$$

where the vectors $\underline{X}$, $\underline{\epsilon}$, $\underline{v}$, and the matrices A, B, C are given by :

$$\underline{X} = (\underline{X}'_1, \underline{X}'_2, ...\underline{X}'_N)',\ \underline{\epsilon} = (\underline{\epsilon}'_1, \underline{\epsilon}'_2, ..., \underline{\epsilon}'_N)' \text{ and } \underline{v} = (\underline{X}'_0, \underline{\epsilon}'_0)',$$

$$A = \begin{pmatrix} I & & & & \\ -\underline{\Phi} & I & & & \\ 0 & -\Phi & I & & \\ & 0 & \ddots & \ddots & \\ 0 & & 0 & -\underline{\Phi} & I \end{pmatrix}_{NS\times NS}, \quad C = \begin{pmatrix} \underline{\Phi} & -\underline{\Theta}_1 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix}_{NS\times 2S}$$

$$\text{and } B^{-1} = \begin{pmatrix} \underline{\Theta}_0 & & & & \\ -\underline{\Theta}_1 & \underline{\Theta}_0 & & & \\ 0 & -\underline{\Theta}_1 & \underline{\Theta}_0 & & \\ & 0 & \ddots & \ddots & \\ 0 & & 0 & -\underline{\Theta}_1 & \underline{\Theta}_0 \end{pmatrix}_{NS\times NS}.$$

Note that the block matrices A and B^{-1} are triangular invertible. The block matrix C is such that only the first block line is non null.

Let $\sigma^2\Gamma = E\underline{X}\underline{X}'$ be the covariance matrix of vector $\underline{X}$. Taking into account that the elements of $\underline{v}$ and of $\underline{\epsilon}$ are uncorrelated random variables, one has then

$$\Gamma = A^{-1}B^{-1}\left(\Sigma + BC\Omega C'B'\right)B'^{-1}A'^{-1}, \tag{2.2.11}$$

where $\Sigma = I_N \otimes D$, $\sigma^2\Omega$ is the covariance matrix of the vector of initial values $\underline{v}$ and, as usual, the symbol $\otimes$ denotes the Kronecker product. Since the elements of $\underline{\epsilon}$ are zero-mean normally distributed, the vector $\underline{X}$ is also normally distributed of zero mean and covariance matrix $\sigma^2\Gamma$. Then the log likelihood function of $\beta = vec\,(\underline{\Phi}, \underline{\Theta}_0, \underline{\Theta}_1, D)$ and σ^2 is given as follows :

$$\log L(\beta, \sigma^2; \underline{X}) = -NS/2\log(2\pi\sigma^2) + \log|\det(\Gamma)|^{-1/2} - \frac{(\underline{X}'\Gamma^{-1}\underline{X})}{2\sigma^2}.$$

As usual, σ^2 may be concentrated out to yield

$$\sigma^2 = \frac{1}{NS}\underline{X}'\Gamma^{-1}\underline{X},$$

so that maximizing $L(\beta, \sigma^2; \underline{X})$ with respect to β and σ^2 is equivalent to minimizing

$$\mathcal{L}(\beta; \underline{X}) = \left(\underline{X}'\Gamma^{-1}\underline{X}\right)(\det(\Gamma))^{1/NS}. \tag{2.2.12}$$

To evaluate this expression, one will base on Dent (1977) and Ljung and Box (1979) which exploited the form of the matrix Γ, from where while serving (2.2.11), the quadratic form in (2.2.12) becomes

$$\underline{X}'\Gamma^{-1}X = \underline{\widetilde{\epsilon}}'\left(\Sigma + BC\Omega C'B'\right)^{-1}\underline{\widetilde{\epsilon}}, \tag{2.2.13}$$

where $\underline{\widetilde{\epsilon}} = BA\underline{X}$ is the vector of $\underline{\epsilon}_i$ which one obtains if one replaces $\underline{v}$ by 0 in the equation (2.2.10). This can be provided by the conditional method

$$\varepsilon_{i+\tau S} = y_{i+\tau S} - \sum_{j=1}^{p_i}\phi_{i,j}y_{i-j+\tau S} + \sum_{j=1}^{q_i}\theta_{i,j}\varepsilon_{i-j+\tau S},$$

for $i = 1, ..., S$ and $\tau = 0, ..., N-1$, with null initial values.

Let us use the fact that BC can be written as B^*C^* where the $(N \times 1)$-block matrix B^* is the first column of B and $C^* = [\underline{\Phi}, -\underline{\Theta}_1]$ is the first (1×2) block-line of C. Thus expression (2.2.13) becomes

$$\underline{X}'\Gamma^{-1}X = \underline{\widetilde{\epsilon}}'\left(\Sigma + B^*C^*\Omega C^{*\prime}B^{*\prime}\right)^{-1}\underline{\widetilde{\epsilon}}. \tag{2.2.14}$$

Otherwise, the well-known Woodbury formula (also known as the matrix inversion lemma MIL, see Chapter 2) states that

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(DA^{-1}B + C^{-1})^{-1}DA^{-1}, \tag{2.2.15}$$

for all matrices A, B, C, V of compatible dimension such that the required inverses exist. Application of (2.2.15) to (2.2.11) with $A = \Sigma$, $B = V = B^*$, $C = C^*\Omega C^{*\prime}$, gives

$$\Gamma^{-1} = A'B'\left[\Sigma^{-1} - \Sigma^{-1}B^*\left[B^{*\prime}B^* + (C^*\Omega C^{*\prime})^{-1}\right]^{-1}B^{*\prime}\right]BA\Sigma^{-1},$$

so that the left member of (2.2.14) will be written

$$\underline{\widetilde{\epsilon}}'\Sigma^{-1}\underline{\widetilde{\epsilon}} - \Sigma^{-1}(B^{*\prime}\underline{\widetilde{\epsilon}})'\left[B^{*\prime}B^* + (C^*\Omega C^{*\prime})^{-1}\right]^{-1}(B^{*\prime}\underline{\widetilde{\epsilon}})\Sigma^{-1},$$

which requires the inverse of a matrix of lower dimension.

The calculation of $\det \Gamma$ in (2.2.12) appears as a sub-product. Using the same notations as in (2.2.15), the calculation of the determinant of a block matrix give

$$\begin{aligned}\det \begin{bmatrix} -A & B \\ D & C^{-1} \end{bmatrix} &= -\det A \det(DA^{-1}B + C^{-1}) \\ &= -\det C^{-1} \det(A + BCD).\end{aligned}$$

It results, since $\det A = 1$ and $\det B = (\det \underline{\Theta}_0)^{-N} = 1$, that

$$\begin{aligned}\det \Gamma &= \left(\Sigma + B^* C^* \Omega C^{*\prime} B^{*\prime}\right) \\ &= \left(\prod_{i=1}^{S} D_i\right)^N \det\left(C^* \Omega C^{*\prime}\right) \det\left(B^{*\prime}\Sigma^{-1}B^* + \left(C^*\Omega C^{*\prime}\right)^{-1}\right),\end{aligned}$$

where D_i are the ith diagonal element of the matrix D. It is easy to check from the construction of the triangular matrix B^{-1} that its inverse, B, has the following structure

$$B = \begin{pmatrix} \underline{\Theta}_0^{-1} & & & & \\ b_{11} & \underline{\Theta}_0^{-1} & & & \\ b_{21} & b_{22} & \underline{\Theta}_0^{-1} & & \\ \vdots & \vdots & & & \\ b_{N-1,1} & b_{N-1,2} & & b_{N-1,N-1} & \underline{\Theta}_0^{-1} \end{pmatrix}$$

from which we need only the first column in order to identify the matrix B^*. This gives, taking into account the fact that $BB^{-1} = I_N \otimes I_S$,

$$\begin{cases} b_{11} = \underline{\Theta}_0^{-1} \underline{\Theta}_1 \underline{\Theta}_0^{-1} \\ \underline{\Theta}_0 b_{j,1} - \underline{\Theta}_1 b_{j-1,1} = 0 \quad j = 2, ..., N-1, \end{cases}$$

or equivalently

$$b_{j,1} = (\underline{\Theta}_1 \underline{\Theta}_0^{-1})^j \underline{\Theta}_0^{-1}, \quad j = 1, ..., N-1 \tag{2.2.16}$$

For higher-period models, exploitation of (1.2.86) for calculating the nth power of $\underline{\Theta}_1$ and $\underline{\Theta}_0^{-1}$ in (2.2.16) can be very useful.

Finally, to evaluate (2.2.12), it remains to determine the covariance matrix $\sigma^2\Omega$ of the vector

$\underline{v}$, which has the following form

$$\begin{aligned}\Omega &= \begin{pmatrix} \sigma^{-2}E\left(\underline{X}_0\underline{X}_0'\right) & \sigma^{-2}E\left(\underline{X}_0\underline{\epsilon}_0'\right) \\ \sigma^{-2}E\left(\underline{\epsilon}_0\underline{X}_0'\right) & \sigma^{-2}E\left(\underline{\epsilon}_0\underline{\epsilon}_0'\right)\end{pmatrix} \\ &= \begin{pmatrix} \Gamma_0 & \Theta_0 D \\ D\Theta_0' & D \end{pmatrix}, \end{aligned} \tag{2.2.17}$$

where $\Gamma_0 = \sigma^{-2}E\left(\underline{X}_0\underline{X}_0'\right)$. The normalized autocovariance Γ_0 is derived by means of the following subsection.

2.2.3 Calculation of the autocovariances for a $VARMA(1,1)$ model

To obtain the theoretical autocovariances of the stable $PARMA$ model (2.2.9a) belonging to the class $\Im_{11}$, we use a procedure based on a modified form of the Yule-Walker equations.

Postmultiplying equation (2.2.9b) by $\underline{X}_{\tau-k}'$ where $k \in \mathbb{N}$, and taking expectations gives the following system

$$\begin{cases} \Gamma_0 - \underline{\Phi}\Gamma_1' = \underline{\Theta}_0 D\underline{\Theta}_0' + \underline{\Theta}_1 D\underline{\Theta}_1' - \underline{\Theta}_1 D\underline{\Theta}_0'\underline{\Phi}' \\ \Gamma_1 - \underline{\Phi}\Gamma_0 = -\underline{\Theta}_1 D\underline{\Theta}_0' \\ \Gamma_k - \underline{\Phi}\Gamma_{k-1} = 0 & k \geq 2 \end{cases}$$

where $\Gamma_k = \sigma^{-2}E\left(\underline{X}_\tau\underline{X}_{\tau-k}'\right)$. The first equation of the latter system of equations can be rewritten as a discrete time algebraic Lyapunov equation ($DTALE$)

$$\begin{cases} \Gamma_0 - \underline{\Phi}\Gamma_0\underline{\Phi}' = \Lambda \\ \Gamma_1 = \underline{\Phi}\Gamma_0 - \underline{\Theta}_1 D\underline{\Theta}_0' \\ \Gamma_k = \underline{\Phi}^k\Gamma_0 - \underline{\Phi}^{k-1}\underline{\Theta}_1 D\underline{\Theta}_0' & k \geq 2, \end{cases} \tag{2.2.18}$$

where $\Lambda = \underline{\Theta}_0 D\underline{\Theta}_0' + \underline{\Theta}_1 D\underline{\Theta}_1' - \underline{\Theta}_1 D\underline{\Theta}_0'\underline{\Phi}' - \underline{\Phi}\underline{\Theta}_0 D\underline{\Theta}_0'$. If Γ_0 and the model parameters are known, the Γ_k can be computed by using (2.2.19). It is worth noting that the use of (1.2.8) for calculating $\underline{\Phi}^k$ in (2.2.18) is very useful in reducing the computational burden. Furthermore, provided that the model parameters are known, Γ_0 can be determined as follows. taking the column stacking operator, *vec*, of both sides of the first equation of (2.2.18) and using the relation $vec(ABC) = (C' \otimes A)\, vec(B)$ we obtain

$$vec\left(\Gamma_0\right) = \left(I_{S^2\times S^2} - \underline{\Phi}\otimes\underline{\Phi}\right)^{-1} vec\left(\Lambda\right), \tag{2.2.19}$$

for a causal $VARMA(1,1)$ model. Moreover, to reduce the computational complexity of (1.4.19), we can apply the *doublin algorithm* (e.g. Anderson and Moore 1979), which allows one to calculate

Γ_0 without explicitly inverting $(I_{S^2\times S^2} - \underline{\Phi} \otimes \underline{\Phi})$. A more efficient technique to obtain Γ_0 is to solve directly the $(DTALE)$ equation

$$\Gamma_0 - \underline{\Phi}\Gamma_0\underline{\Phi}' = \Lambda,$$

using standard methods (Barraud, 1977; Kitagawa, 1977; Hammarling, 1982).

Finally to get the autocovariances $\gamma_k^{(i)} = E\left(y_{i+S\tau} y_{i+S\tau-k}\right)$ of the periodic model (2.2.9a), for $i = 1, ..., S$ and $k \in \mathbb{Z}$, we solve the system (2.2.19) on Γ_k and use the relation

$$\Gamma_k(i,j) = \sigma^{-2}\gamma_{i-j+Sk}^{(i)},$$

with $\gamma_{-k}^{(i)} = \gamma_k^{(i+k)}$ and $\gamma_k^{(i+S\tau)} = \gamma_k^{(i)}$.

2.3 Asymptotic inference for $PVARMA$ models

It is well known that the maximum likelihood, the weighted least squares and the Yule-Walker methods provide asymptotically equivalent estimators for PAR models. In this simple case, the asymptotic properties can be found in Pagano (1978), which has studied the Yule-Walker estimates for PAR models. For the general $PARMA$ case, large statistical properties of the least squares and the maximum likelihood estimates have been considered by Basawa and Lund (2001). In this section we give a brief survey of some results concerning the large sample properties of parameter estimates for $PVARMA$ models. The main results given here generalize those given in Basawa and Lund (2001) for the univariate $PARMA$ case. Thus, we omit the proofs and present only a brief discussion with related references for each result. The main purpose here is to provide the necessary background for the next chapters, concerning the asymptotic properties of on-line estimates for $PARMA$ models (Chapter 3) and periodic $GARCH$ models (Chapter 5). Thus, we concentrate in our study on the class of methods that directly use the data (e.g. the maximum likelihood and least squares methods) because they generally provide asymptotically efficient estimates, even in the presence of a moving average component. The first subsection is devoted to analyzing the properties of the (ordinary and weighted) least squares estimates, while the second treats the large ML estimates properties.

Throughout the rest of this chapter we use β to denote the parameter of a model. Occasionally, when analyzing the (limiting) statistical properties of different estimates, we shall use β^* to denote the true value of the model parameters, thus assuming that the observed data have actually been generated from a particular model. This value β^* is, of course, available only to the analyst and not to the user. However, we use β^* and β interchangeably to represent the model parameters, the latter notation being preferred when emphasis on the true value is not paramount.

2.3.1 Least squares estimation of $PARMA$ models

Consider a $PVARMA$ model given by the following stochastic difference equation

$$\mathbf{y}_t - \sum_{j=1}^{p} \boldsymbol{\phi}_{tj}\, \mathbf{y}_{t-j} = \boldsymbol{\varepsilon}_t - \sum_{j=1}^{q} \boldsymbol{\theta}_{tj}\, \boldsymbol{\varepsilon}_{t-j},\ t \in \mathbb{Z}, \tag{2.3.1}$$

where $\{\boldsymbol{\varepsilon}_t, t \in \mathbb{Z}\}$ is an m-variate strongly periodic white noise (not necessarily Gaussian) with covariance matrix Σ_t. The independence assumption on $\{\boldsymbol{\varepsilon}_t\}$ can be replaced by a less stringent assumption, which requires this sequence to be only a multivariate Martingale difference sequence with periodic covariance $E\left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' / \mathcal{F}_{t-1}\right) = \Sigma_t$, where $\mathcal{F}_t$ denotes the $\sigma-$Algebra generated by $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, ..., \boldsymbol{\varepsilon}_t$. Suppose that model (2.3.1) is causal and invertible (i.e. the assumption $A2$ given in Section 1.4 holds) and that the following fourth moment assumption holds.

$A3$: The periodically ergodic process $\{\mathbf{y}_t;\ t \in \mathbb{Z}\}$ is such that $E\left(\mathbf{y}_t^{\otimes 4}\right) < \infty$, for any $t \in \mathbb{Z}$.

Let $\beta = \left(vec\left(\boldsymbol{\phi}\right)', vec\left(\boldsymbol{\theta}\right)'\right)'$ be the autoregressive and moving average parameters, where $\boldsymbol{\phi} = (\boldsymbol{\phi}_1, ..., \boldsymbol{\phi}_p)$ and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_q)$ with $\boldsymbol{\phi}_i = (\boldsymbol{\phi}_{i,1}, \boldsymbol{\phi}_{i,2} ..., \boldsymbol{\phi}_{i,p})$ and $\boldsymbol{\theta}_i = (\boldsymbol{\theta}_{i,1}, \boldsymbol{\theta}_{i,2} ..., \boldsymbol{\theta}_{i,q})$, $i = 1, 2, ..., S$. The periodic white noise covariance matrices are denoted by $\Sigma = (\Sigma_1, \Sigma_2, ..., \Sigma_S)$.

Suppose that we have a data sample $\{\mathbf{y}_1, ..., \mathbf{y}_{NS}\}$ generated from model (2.3.1), which we rewrite as

$$\mathbf{y}_t - \sum_{j=1}^{p} \boldsymbol{\phi}_{tj}^*\, \mathbf{y}_{t-j} = \boldsymbol{\varepsilon}_t - \sum_{j=1}^{q} \boldsymbol{\theta}_{tj}^*\, \boldsymbol{\varepsilon}_{t-j},\ t \in \mathbb{Z},$$

where $\beta^* = \left(vec\left(\boldsymbol{\phi}^*\right)', vec\left(\boldsymbol{\theta}^*\right)'\right)'$ is the true value of the unknown parameter vector β. Let

$$\boldsymbol{\varepsilon}_t(\beta) = \mathbf{y}_t - \widetilde{\mathbf{y}}_t(\beta)$$

be the prediction error function obtained when approximating $\mathbf{y}_t$ by a prediction $\widetilde{\mathbf{y}}_t(\beta)$, which is obtained from

$$\widetilde{\mathbf{y}}_t(\beta) = \sum_{j=1}^{p} \boldsymbol{\phi}_{tj} \mathbf{y}_{t-j} - \sum_{j=1}^{q} \boldsymbol{\theta}_{tj} \boldsymbol{\varepsilon}_{t-j}\left(\beta\right),\ t \in \mathbb{Z}, \tag{2.3.2}$$

for a given β. Assuming $\boldsymbol{\varepsilon}_t(\beta^*) = \boldsymbol{\varepsilon}_t$, it is clear that $\widetilde{\mathbf{y}}_t(\beta^*) = E(\mathbf{y}_t/\mathcal{F}_{t-1})$ represents the best prediction of $\mathbf{y}_t$ (in the mean squares sense) based on data until time $t-1$, where $\mathcal{F}_t$ denotes the $\sigma-$Algebra generated by $\varepsilon_1, \varepsilon_2, ..., \varepsilon_t$.

Ordinary least squares

The ordinary least squares estimate, $\widetilde{\beta}_{LS}$, of β, based on a data sample $\{\mathbf{y}_1, ..., \mathbf{y}_{NS}\}$, is an argument of β that minimizes

$$\widetilde{V}_{LS}(\beta) = \frac{1}{NS} \sum_{k=0}^{N-1} \sum_{i=1}^{S} \left(\mathbf{y}_{i+Sk} - \widetilde{\mathbf{y}}_{i+Sk}(\beta)\right)' \left(\mathbf{y}_{i+Sk} - \widetilde{\mathbf{y}}_{i+Sk}(\beta)\right), \tag{2.3.3}$$

The prediction function $\widetilde{\mathbf{y}}_t(\beta)$ can be obtained from (2.3.2). Furthermore, under the Gaussian assumption on the innovations, such a prediction can be computed recursively from the Kalman recursion (see Section 2.2) or by means of the innovations algorithm (cf, Anderson et al., 1999; Lund and Basawa, 2000).

As with classical $ARMA$ analyses, several approximations are made in order to simplify the analysis. To compute $\boldsymbol{\varepsilon}_t(\beta)$ from the data sample, one can use the invertible representation

$$\boldsymbol{\varepsilon}_t(\beta) = \sum_{j=0}^{\infty} \boldsymbol{\pi}_{t,j} \mathbf{y}_{t-j}, \tag{2.3.4}$$

under the invertibility condition (on β), which implies that

$$\sum_{j=0}^{\infty} \|\boldsymbol{\pi}_{t,j}\| < \infty,$$

for all $t \in \{1, ..., S\}$, where the periodic coefficients $(\boldsymbol{\pi}_{t,j})$ can be obtained from the model parameters as follows (see, e.g. Lund and Basawa (2000) for the univariate case).

$$\begin{cases} \boldsymbol{\pi}_{s,0} = \mathbf{I}_m, \\ \boldsymbol{\pi}_{s,k} = -\boldsymbol{\phi}_{s,k} 1_{[k \leq q]} + \sum\limits_{j=1}^{\min(k,p)} \boldsymbol{\theta}_{s,j} \boldsymbol{\pi}_{s-j,k-j}, \end{cases} \quad s = 1, ..., S.$$

Unfortunately the computation of $\boldsymbol{\varepsilon}_t(\beta)$ via (2.3.4) is not possible as $\mathbf{y}_0, \mathbf{y}_{-1}, ...$ are unobservable. However the sequence $\{\boldsymbol{\varepsilon}_t(\beta)\}$ can be approximated by the sequence $\{\widehat{\boldsymbol{\varepsilon}}_t(\beta)\}$ which satisfies the following conditional stochastic difference equation (a version of the $PVARMA$ model)

$$\mathbf{y}_t - \sum_{j=1}^{p} \boldsymbol{\phi}_{tj}\, \mathbf{y}_{t-j} = \widehat{\boldsymbol{\varepsilon}}_t(\beta) - \sum_{j=1}^{q} \boldsymbol{\theta}_{tj}\, \widehat{\boldsymbol{\varepsilon}}_{t-j}(\beta),\ t \geq 1, \tag{2.3.5}$$

with $\widehat{\boldsymbol{\varepsilon}}_t(\beta) = \mathbf{y}_t = \mathbf{0}$ for $t \leq 0$. Similar techniques to the $ARMA$ case (see e.g. Brockwell and Davis, 1991, Section 8.11) show that $\widehat{\boldsymbol{\varepsilon}}_t(\beta)$ can be given by

$$\widehat{\boldsymbol{\varepsilon}}_t(\beta) = \sum_{j=0}^{t-1} \boldsymbol{\pi}_{t,j}\mathbf{y}_{t-j}.$$

Using the approximation $\widehat{\boldsymbol{\varepsilon}}_t(\beta)$ for $\widetilde{\boldsymbol{\varepsilon}}_t(\beta)$ in (2.3.3) gives an approximate least squares function criterion, which will be denoted by $\widehat{V}_{LS}\left(\beta\right)$:

$$\widehat{V}_{LS}\left(\beta\right) = \frac{1}{NS} \sum_{k=0}^{N-1} \sum_{i=1}^{S} \widehat{\boldsymbol{\varepsilon}}_{i+Sk}(\beta)' \widehat{\boldsymbol{\varepsilon}}_{i+Sk}(\beta). \tag{2.3.6}$$

Let $\widehat{\beta}_{LS}$ be an argument that minimizes $\widehat{V}_{LS}\left(\beta\right)$. It is easy to show that $\widehat{\beta}_{LS}$ is a solution of the normal equation

$$\sum_{k=0}^{N-1} \sum_{i=1}^{S} \left(\frac{\partial \widehat{\boldsymbol{\varepsilon}}_{i+Sk}(\beta)'}{\partial \beta} \right) \widehat{\boldsymbol{\varepsilon}}_{i+Sk}(\beta) = \mathbf{0}_{(p+q)m \times 1}. \tag{2.3.7}$$

Similar arguments as that in Section 8.11 of Brockwell and Davis (1991) show that $\widehat{\beta}_{LS}$ and $\widetilde{\beta}_{LS}$ have the same limiting distribution (see Basawa and Lund (2001) for the univariate $PARMA$ case). The adaptation to the multivariate $PVARMA$ case is tedious but straightforward. Therefore, we only study the properties of the estimator $\widehat{\beta}_{LS}$.

Before presenting the limiting distribution of $\widehat{\beta}_{LS}$, we need some convergence results concerning a version of the law of large numbers. For simplicity consider the following notations.

Put $Z_k\left(\beta\right) = -\sum_{i=1}^{S} \psi_{i+Sk}\left(\beta\right) \boldsymbol{\varepsilon}_{i+Sk}\left(\beta\right)$, $S_N\left(\beta\right) = \sum_{k=0}^{N-1} Z_k\left(\beta\right)$ and let $\psi_t\left(\beta\right) = -\dfrac{\partial \boldsymbol{\varepsilon}_t(\beta)'}{\partial \beta}$ denote the $m(p+q) \times m$ gradient matrix. The following lemma shows that the gradient matrix $\psi_t\left(\beta^*\right)$ satisfies a law of large numbers.

Lemma 2.3.1

i) *Under the above assumptions, we have*

$$\frac{1}{N}\sum_{k=0}^{N-1}\psi_{i+Sk}\left(\beta^*\right)\xrightarrow{p}\mathbf{0}$$

and

$$\frac{1}{N}\sum_{k=0}^{N-1}\psi_{i+Sk}\left(\beta^*\right)\psi'_{i+Sk}\left(\beta^*\right)\xrightarrow{p}E\left[\psi_{i+Sk}\left(\beta^*\right)\psi'_{i+Sk}\left(\beta^*\right)\right]$$

as $N\rightarrow\infty$ *for* $i\in\{1,...,S\}$.

ii) Moreover,

$$\frac{1}{N}S_N\left(\beta^*\right)\xrightarrow{p}0$$

$$\frac{1}{N}\sum_{k=0}^{N-1}Z_k\left(\beta^*\right)Z_k\left(\beta^*\right)'\xrightarrow{p}F_1\left(\beta^*,\Sigma^*\right)$$

$$\frac{1}{N}\left.\frac{\partial S_N\left(\beta\right)}{\partial\beta}\right|_{\beta=\beta^*}\xrightarrow{p}F_2\left(\beta^*\right)$$

where

$$\begin{aligned} F_1\left(\beta^*,\Sigma\right) &= \sum_{i=1}^{S}E\left[\psi_{i+Sk}\left(\beta^*\right)\Sigma_i\psi'_{i+Sk}\left(\beta^*\right)\right] \\ &\text{and} \\ F_2\left(\beta^*\right) &= \sum_{i=1}^{S}E\left[\psi_{i+Sk}\left(\beta^*\right)\psi'_{i+Sk}\left(\beta^*\right)\right] \end{aligned} \tag{2.3.8}$$

Proof

The proof is a straightforward adaptation of that given in Basawa and Lund (2001, Section 4) to the multivariate case, hence it will be omitted. ■

The following theorem shows the $\sqrt{N}-$consistency of the least squares estimates $\widehat{\beta}_{LS}$ and $\widetilde{\beta}_{LS}$.

Theorem 2.3.2

Under the above assumptions, we have

$$\sqrt{N}\left(\widehat{\beta}_{LS}-\beta^*\right)\underset{N\longrightarrow\infty}{\xrightarrow{D}}\mathcal{N}\left(0,F_2^{-1}\left(\beta^*\right)F_1\left(\beta^*,\Sigma^*\right)F_2^{-1}\left(\beta^*\right)\right) \tag{2.3.9}$$

where Σ^* *is the true value of the parameter covariances* Σ^*.

Proof

The proof is classical in the literature. It is based on the Taylor expansion of $S_N\left(\widehat{\beta}_{LS}\right)$ on a neighborhood of β^*, a central limit version for martingales (e.g. Hall and Heyde, 1980, Chapter 3) and Lemma 1.6.1. See Basawa and Lund (2001) for the univariate case.■

It is worth noting that (2.3.7) does not involve the parameter covariances Σ, therefore, the least squares estimator can be obtained without any information about Σ^*.

Let $\widehat{\Sigma}_i^{(LS)}$ be the ordinary least squares estimator of Σ_i^* by averaging product errors during season i, $i = 1, ..., S$ (cf, Basawa and Lund 2001, Adams and Goodwin 1995 for the univariate $PARMA$)

$$\widehat{\Sigma}_i^{(LS)} = \frac{1}{N}\sum_{k=0}^{N-1}\widehat{\boldsymbol{\varepsilon}}_{i+Sk}(\widehat{\beta}_{LS})\widehat{\boldsymbol{\varepsilon}}_{i+Sk}(\widehat{\beta}_{LS})' \tag{2.3.10}$$

The following theorem assures the consistency and asymptotic normality of $\widehat{\Sigma}_i^{(LS)}$.

Theorem 2.3.2

Under the above assumptions, we have

$$\sqrt{N}\left(vech\left(\widehat{\Sigma}_i^{(LS)}\right) - vech\left(\Sigma_i^*\right)\right) \underset{N\longrightarrow\infty}{\overset{D}{\longrightarrow}} \mathcal{N}\left(0, E\left(\Xi_i\Xi_i'\right)\right) \tag{2.3.11}$$

for any $i = 1, ..., S$, *where*

$$\Xi_i = vech\left(\boldsymbol{\varepsilon}_{i+Sk}\boldsymbol{\varepsilon}_{i+Sk}' - E\left(\boldsymbol{\varepsilon}_{i+Sk}\boldsymbol{\varepsilon}_{i+Sk}'\right)\right)$$

and $vech(A)$ *denotes the half column stacking of the matrix* A.

Proof

Standard arguments show that the asymptotic distribution of

$$\frac{1}{N}\sum_{k=0}^{N-1}\boldsymbol{\varepsilon}_{i+Sk}(\widehat{\beta}_{LS})\boldsymbol{\varepsilon}_{i+Sk}(\widehat{\beta}_{LS})'$$

and that of

$$\frac{1}{N}\sum_{k=0}^{N-1}\widehat{\boldsymbol{\varepsilon}}_{i+Sk}(\widehat{\beta}_{LS})\widehat{\boldsymbol{\varepsilon}}_{i+Sk}(\widehat{\beta}_{LS})'$$

are identical. Moreover, application of the central limit theorem for the latter expression gives

$$\frac{1}{\sqrt{N}}\left(vech\left(\sum_{k=0}^{N-1}\boldsymbol{\varepsilon}_{i+Sk}(\widehat{\beta}_{LS})\boldsymbol{\varepsilon}_{i+Sk}(\widehat{\beta}_{LS})'\right)-vech\left(\Sigma_i^*\right)\right)\underset{N\longrightarrow\infty}{\overset{D}{\longrightarrow}}\mathcal{N}\left(0,E\left(\Xi_i\Xi_i'\right)\right),$$

for any $i=1,...,S$. Combining these two facts achieve the proof. ■

It should be noted that under the independence assumption on $\{\boldsymbol{\varepsilon}_t\}$, Theorem 2.3.2 means that $\widehat{\Sigma}_i^{(LS)}$, $i=1,...,S$ are asymptotically independent.

Weighted least squares

The weighted least squares estimate of β, denoted by $\widetilde{\beta}_{WLS}$, based on a data sample $\mathbf{y}_1,...,\mathbf{y}_{NS}$ is an argument of β that minimizes the weighted sum of squares

$$\widetilde{V}_{WLS}\left(\beta,\Sigma\right)=\frac{1}{NS}\sum_{k=0}^{N-1}\sum_{i=1}^{S}\left(\mathbf{y}_{i+Sk}-\widetilde{\mathbf{y}}_{i+Sk}\left(\beta\right)\right)'h_{i+Sk}^{-1}(\beta,\Sigma)\left(\mathbf{y}_{i+Sk}-\widetilde{\mathbf{y}}_{i+Sk}\left(\beta\right)\right),\qquad(2.3.12)$$

where as in Section 2.2

$$h_t(\beta,\Sigma)=E\left[\left(\mathbf{y}_t-\widetilde{\mathbf{y}}_t\left(\beta\right)\right)\left(\mathbf{y}_t-\widetilde{\mathbf{y}}_t\left(\beta\right)\right)'/\,\mathbf{y}_1,...,\mathbf{y}_{t-1}\right],$$

is the conditional prediction error variance, which coincides, for the Gaussian assumption on the innovations model, with the unconditional ones

$$E\left[\left(\mathbf{y}_t-\widetilde{\mathbf{y}}_t\left(\beta\right)\right)'\left(\mathbf{y}_t-\widetilde{\mathbf{y}}_t\left(\beta\right)\right)\right],$$

(see Brockwell and Davis 1991, Section 2.3).

As in the ordinary least squares analysis, the limiting properties of the estimates are the same when one minimizes the approximated weighted sum of squares

$$\widehat{V}_{WLS}\left(\beta,\Sigma\right)=\frac{1}{NS}\sum_{k=0}^{N-1}\sum_{i=1}^{S}\widehat{\boldsymbol{\varepsilon}}_{i+Sk}(\beta)'\Sigma_i^{-1}\widehat{\boldsymbol{\varepsilon}}_{i+Sk}(\beta),\qquad(2.3.13)$$

in which Σ_i and $\widehat{\boldsymbol{\varepsilon}}_{i+Sk}(\beta)$ approximate $h_{i+Sk}(\beta,\Sigma)$ and $\mathbf{y}_{i+Sk}-\widetilde{\mathbf{y}}_{i+Sk}\left(\beta\right)$, respectively. Note that approximating $h_{i+Sk}(\beta,\Sigma)$ by Σ_i is quite natural since one can easily show that

$$h_{i+Sk}(\beta,\Sigma)\overset{k\to\infty}{\to}\Sigma_i,$$

(for the univariate case with the sample innovations obtained from the innovations algorithm see part (i) of Proposition 3.1 in Lund an Basawa (2000)).

Let $\widehat{\beta}_{WLS}$ be an argument that minimizes $\widehat{V}_{WLS}(\beta, \Sigma)$, for a fixed Σ. It is easy to show that $\widehat{\beta}_{WLS}$ is a solution of the estimating equation

$$\sum_{k=0}^{N-1}\sum_{i=1}^{S}\left(\frac{\partial\widehat{\varepsilon}_{i+Sk}(\beta)}{\partial\beta}\right)\Sigma_i^{-1}\widehat{\varepsilon}_{i+Sk}(\beta) = \mathbf{0}_{m(p+q)\times 1}. \tag{2.3.14}$$

As in the ordinary least squares case, to find the asymptotic distribution of the parameter estimate $\widehat{\beta}_{WLS}$ we need similar convergence results as Lemma 2.3.1. Consider the following notations

$$Z_k^{(W)}(\beta, \Sigma) \stackrel{def}{=} \sum_{i=1}^{S}\psi_{i+Sk}(\beta)\,\Sigma_i^{-1}\varepsilon_{i+Sk}(\beta),$$

and

$$S_N^{(W)}(\beta, \Sigma) \stackrel{def}{=} \sum_{k=0}^{N-1} Z_k^{(W)}(\beta, \Sigma).$$

Then we can obtain the following result

Lemma 2.3.2

Under the above assumptions, we have

$$\begin{aligned}
&\frac{1}{N}S_N^{(W)}(\beta^*) \xrightarrow{p} 0 \\
&\frac{1}{N}\sum_{k=0}^{N-1} Z_k^{(W)}(\beta^*)\,Z_k^{(W)}(\beta^*)' \xrightarrow{p} F(\beta^*, \Sigma^*) \\
&\frac{1}{N}\left.\frac{\partial S_N^{(W)}(\beta)}{\partial\beta}\right|_{\beta=\beta^*} \xrightarrow{p} F(\beta^*, \Sigma^*)
\end{aligned} \tag{2.3.15}$$

where $F(\beta, \Sigma) = \sum_{i=1}^{S} E\left[\psi_{i+Sk}(\beta)'\,\Sigma_i^{-1}\psi_{i+Sk}(\beta)\right]$ *denotes the asymptotic Fisher information matrix (as given in Section* 1.4. for the univariate $PARMA$ case).

Proof

See for example Basawa and Lund (2001). ■

The following theorem show the asymptotic efficiency of the WLS-estimator $\widehat{\beta}_{WLS}$.

Theorem 2.3.3

Under the above assumptions, we have

$$\sqrt{N}\left(\widehat{\beta}_{WLS}-\beta^*\right) \underset{N\longrightarrow\infty}{\overset{D}{\longrightarrow}} \mathcal{N}\left(0, F^{-1}\left(\beta^*, \Sigma^*\right)\right) \tag{2.3.16}$$

Proof

The proof is similar to that of Theorem 2.3.1. ■

2.3.2 Maximum likelihood estimation of $PVARMA$ models

As shown in Section 2.2 (see (2.2.3)), the Gaussian likelihood of the $PVARMA$ model (2.3.1), for a sample data $\mathbf{y}$ of size NS, is given by

$$\mathcal{L}(\beta, \Sigma; \mathbf{y}) = (2\pi)^{-NS/2} \prod_{k=0}^{N-1}\prod_{i=1}^{S} (\det(h_{i+Sk}))^{-1/2} \exp\left\{-\frac{1}{2}\sum_{k=0}^{N-1}\sum_{i=1}^{S} \widehat{\boldsymbol{\varepsilon}}'_{i+Sk}\mathbf{h}^{-1}_{i+Sk}\widehat{\boldsymbol{\varepsilon}}_{i+Sk}\right\}, \tag{2.3.17}$$

where $\widehat{\boldsymbol{\varepsilon}}_{i+Sk}$ is the innovation sample obtained from the Kalman recursion (2.2.6).

As in the least squares analysis, using the approximation Σ_i for $\mathbf{h}_{i+Sk}$ in (2.3.17) gives an approximate likelihood function which will be denoted by $\widehat{\mathcal{L}}$

$$\widehat{\mathcal{L}}(\beta, \Sigma; \mathbf{y}) = (2\pi)^{-NS/2} \prod_{i=1}^{S} (\det(\Sigma_i))^{-N/2} \exp\left\{-\frac{1}{2}\sum_{k=0}^{N-1}\sum_{i=1}^{S} \widehat{\boldsymbol{\varepsilon}}'_{i+Sk}\left(\beta\right)\Sigma_i^{-1}\widehat{\boldsymbol{\varepsilon}}_{i+Sk}\left(\beta\right)\right\} \tag{2.3.18}$$

As is well known, the estimators obtained by maximizing $\mathcal{L}$ and $\widehat{\mathcal{L}}$, respectively, are asymptotically equivalent. Therefore, we only study the estimator obtained when maximizing (2.3.18). Let $\widehat{\beta}_{ML}$ and $\widehat{\Sigma}_i^{ML}$, $i = 1, ..., S$ be arguments that maximize $\widehat{\mathcal{L}}(\beta, \Sigma; \mathbf{y})$. Then, differentiating (2.3.18) with respect to β and Σ gives the following normal equations

$$\sum_{k=0}^{N-1}\sum_{i=1}^{S}\left(\frac{\partial\widehat{\varepsilon}_{i+Sk}(\beta)}{\partial\beta}\right)\Sigma_i^{-1}\widehat{\varepsilon}_{i+Sk}(\beta) = \mathbf{0}_{m(p+q)\times 1} \tag{2.3.19a}$$

$$\frac{1}{N}\sum_{k=0}^{N-1}\widehat{\boldsymbol{\varepsilon}}_{i+Sk}(\beta)\widehat{\boldsymbol{\varepsilon}}_{i+Sk}(\beta)' - \Sigma_i = \mathbf{0}_{m\times m} \tag{2.3.19b}$$

Combining (2.3.19a) and (2.3.19b) shows that $\widehat{\beta}_{ML}$ and $\widehat{\Sigma}_i^{ML}$ are the solutions to

$$\sum_{k=0}^{N-1}\sum_{i=1}^{S}\left(\frac{\partial\widehat{\varepsilon}_{i+Sk}(\beta)}{\partial\beta}\right)\Bigg|_{\beta=\widetilde{\beta}_{ML}}\left(\frac{1}{N}\sum_{k=0}^{N-1}\widehat{\varepsilon}_{i+Sk}(\widehat{\beta}_{ML})\widehat{\varepsilon}_{i+Sk}(\widehat{\beta}_{ML})'\right)^{-1} \times\widehat{\varepsilon}_{i+Sk}(\widehat{\beta}_{ML})=\mathbf{0}_{m(p+q)\times 1} \tag{2.3.20a}$$

$$\widehat{\Sigma}_i^{ML}=\frac{1}{N}\sum_{k=0}^{N-1}\widehat{\boldsymbol{\varepsilon}}_{i+Sk}(\widehat{\beta}_{ML})\widehat{\boldsymbol{\varepsilon}}_{i+Sk}(\widehat{\beta}_{ML})\ ,\ \ i=1,...,S \tag{2.3.20b}$$

Clearly, these last equations show that the ML and the WLS estimators are asymptotically equivalent. The following theorem shows the consistency and the asymptotic normality of the ML estimator $\widetilde{\beta}_{ML}$.

Theorem 2.3.4

Under the above assumptions, we have

$$\sqrt{N}\left(\widehat{\beta}_{ML}-\beta^*\right)\underset{N\longrightarrow\infty}{\overset{D}{\longrightarrow}}\mathcal{N}\left(0,F^{-1}\left(\beta^*,\Sigma^*\right)\right) \tag{2.3.21}$$

where $F\left(\beta^*,\Sigma^*\right)$ *is the asymptotic Fisher information matrix as given above.*

Proof

See Theorem 2.3.3■

2.4 Summary

In this chapter, the off-line estimation problem of periodic $VARMA$ models has been studied. In the first part, an efficient calculation of the $PVARMA$ likelihood, by means of the Kalman filter, has been proposed. In the univariate case, the resulting algorithm differs from existing algorithms (e.g. Jymenez et al., 1989) by the starting Kalman recursion, which is the key of our algorithm efficiency. The proposed likelihood evaluation for $PVARMA$ models is no more complex than the evaluation of the stationary $VARMA$ likelihood. However, some useful devices such as the Chandrasekhar type recursions (e.g. Morf et al., 1974; Houacine and Demoment, 1986) which have been generalized to the periodic case by Aknouche and Hamdi (2006), are needed to improve Kalman recursions for periodic linear systems. In the second part of the chapter the large sample properties of parameter estimates for $PVARMA$ models have been investigated. We have shown that the class of methods that use data directly provides consistent, asymptotically Gaussian and, in some cases, asymptotically efficient estimators. The main limitation of these estimation methods is their computational complexities and flat character (especially for the ML method) caused by the large number of parameters involved by the $PVARMA$ formulation. Moreover, these methods do not want to fit the model from scratch every time a new data point is taken. The following chapter treats this problem.

Chapter 3

On-line estimation of periodic linear models

3.1 Introduction

Parameter estimation problems for time series models have seen significant recent interest from statisticians and engineers. However, most of the existing estimation methods and algorithms are mainly concerned with classical time-invariant time series models. Moreover, it appears that the quasi-totality of existing procedures are of an off-line nature which deal with a beforehand fixed size time series. The need to deal with a progressive size is becoming, nowadays, increasingly necessary. Indeed, in many time series application fields— such as transmission, telecommunications, signal treatment, meteorology, financial bourses and many others — the need to deal with variable series sizes is clear. In short, off-line estimation methods are inefficient when used in the presence of on-line data, mainly because their applications require, for each new available observation, start-from-scratch computations. Off-line estimation algorithms are very costly in terms of the spatial and temporal complexities. For all these reasons, on-line estimation methods are appealing (or the recursive identification methods, as they are known in the engineering literature, e.g. Kushner and Clark, 1978; Ljung and Söderström, 1983; Goodwin and Sin, 1984; Young, 1984; Benveniste et al., 1990; Solo and Kong 1995; Kushner and Yin, 1997; Ljung, 1999; Pollock, 2003 and further references therein), to reduce the computation time and save memory space.

The main usefulness of the recursive form of on-line algorithms is due to the possibility of processing data in real time because of a finite memorizing in a vector of fixed size. Indeed the current estimate is updated at each new available observation, so that the number of operations and the memory space do not increase with respect to the sample size, and depend only on the selected model.

Existing recursive estimation algorithms in the literature are usually associated with a particular model parametrization and the approach by which the method is built. Several approaches can be taken to derive a recursive algorithm. We mention in particular the *off-line approach* for on-line algorithms, the *Bayesian approach*, the *stochastic approximation* approach, and the *pseudo regression* method. These four approaches cover the derivation of most suggested methods (see e.g. Ljung and Söderström (1983) which contains extensive examples on this subject). However, the resulting algorithms exhibit very similar features and can be interpreted in different ways. Another way of expressing this is to say that a given recursive algorithm can be obtained from different approaches.

On-line parameter estimation methods for $PARMA$ models are still, despite their remarkable simplicity, much less common than off-line methods. Indeed, it appears that the work of Adams and Goodwin (1995) is the first one to focus on this interesting problem. Adams and Goodwin (1995) adapted recursive pseudo-linear regression methods to estimate the parameters of $PARMA$ models. Moreover, they studied, based on the works of Solo (1979) and Goodwin and Sin (1984), the convergence problem of their algorithms. Unfortunately, the obtained estimators are not efficient since they are based on the ordinary least squares estimation methods. Moreover, no information about their asymptotic distribution is given.

In this chapter, we mainly provide an adaptation of some on-line recursive algorithms, well-known in the literature of time-invariant $ARMA$ models, to deal with $PARMA$ models. Though transforming $PARMA$ estimation problems to the ones for multivariate $VARMA$ models where several efficient on-line methods are available, the large number of parameters in the corresponding $VARMA$ models suggests direct development of a specific $PARMA$ theory would be fruitful (see Pagano, 1978; Basawa and Lund, 2001).

The rest of this chapter is organized as follows. In Section 2, we review some basic principles and criteria of recursive on-line estimation. Section 3 concentrates on the on-line estimation problem for PAR models. Furthermore, we provide an adaptation of the well-known recursive least squares (RLS) algorithm to deal with PAR models. We show under mild conditions that the obtained estimators are asymptotically efficient. The following sections are then concerned with the general $PARMA$ recursive estimation. In particular, Section 4 generalizes the obtained $PRLS$ algorithm to the $PARMA$ case. The generalization is done by means of the pseudo-linear regression approach (PLR). Several variants of the PLR method (such as the extended least squares, ELS) are developed to take into account the stochastic convergence problem. The developed $PELS$ (Periodic ELS) algorithms provide strongly consistent estimators. In Section 5, we explore an alternative approach based on the recursive predictor error method ($RPEM$). In particular, we extend the well-known recursive maximum likelihood (RML) algorithm to the $PARMA$ case. The obtained estimates are shown to be asymptotically efficient. In Section 6, we show that both $PRML$ (Periodic RML) and $PELS$ algorithms can be formulated via a single set of recursions through a certain prefilter. Moreover, it will be shown that with a modified version of this prefilter, the resulting algorithm lies somewhere in between the $PELS$ an the $PRML$. Section 7 develops an improved variant of the $PRML$ algorithm based on the Fisher information matrix. The resulting algorithm is an on-line analogue of the well-known Fisher scoring algorithm, which is devoted exclusively to the off-line case. Finally, the finite properties of the obtained estimates are studied in the last section by means of a simulation study. Furthermore, we show, through various applications of this algorithm on many $PARMA$ models, that various desirable properties take place in our periodic case.

3.2 Principle and criteria of recursive estimation

Next, we shall recall the general principle of a recursive estimation problem.

3.2.1 Principle

Consider a second order stochastic process $\{y_t,\, t \in \mathbb{Z}\}$, which is supposed to satisfy a certain stochastic model, parametrized in terms of a parameter vector β; and let $y^t = \{y_1, y_2, \ldots, y_t\}$ be the available realization of the underlying process at time t. Then, based on the observation y^t, a recursive estimates of β, denoted by $\widehat{\beta}_t = \widehat{\beta}\left(t, y^t\right)$, is a mapping from the data to the model parameters, subject to the special restriction that *the temporal and the spatial complexities must be independent of the sample size*, i.e., the memory space and the computing time involved in the calculation of $\widehat{\beta}_t$ do not increase with t. Basically, the idea of a recursive algorithm is to condense the information contained in y^t into an estimate $\widehat{\beta}_t$ and in an auxiliary quantity Q_t, and then updating $\widehat{\beta}_t$ and Q_t from their previous value at time $t-1$ (namely $\widehat{\beta}_{t-1}$ and Q_{t-1}) and from the new observation y_t. The recursive estimate $\widehat{\beta}_t$ and the quantity Q_t will be calculated according to an algorithm of the following structure (cf, Ljung and Söderström 1983)

$$\begin{cases} \widehat{\beta}_t = F_t(\widehat{\beta}_{t-1}, Q_t, y_t) \\ Q_t = H_t(Q_{t-1}, \widehat{\beta}_{t-1}, y_t) \end{cases} \tag{3.2.1}$$

where F_t and H_t are functions depending on the used method and the selected model (generally we have $F_t = F$ and $H_t = H$, $\forall t$). Thus, the recursive estimate $\widehat{\beta}_t$ is calculated for each t, from current data, the previous estimate and the auxiliary variable Q_t. These quantities are updated with a fixed number of operations that does not depend on time t. The problem of recursive estimation therefore is reduced to determine the functions (F_t) and (H_t) and in general to analyze the statistical properties of the obtained estimates.

3.2.2 Criteria

Suppose the model that generates the data y^t is given by

$$y_t = f_t\left(y_{t-1}, y_{t-2}, ...y_{t-p}, \varepsilon_{t-1}, \varepsilon_{t-2}, ...\varepsilon_{t-q}; \beta^*\right) + \varepsilon_t, \ \ t \in \mathbb{Z}$$

where $f_t(., ..., .)$ is a measurable function, $\{\varepsilon_t,\, t \in \mathbb{Z}\}$ is an innovation process satisfying certain suitable conditions, and β^* is the true value of the unknown parameter vector. Let $\varepsilon_t(\beta) = y_t - y_t(\beta)$ be the prediction error function obtained when approximating y_t by a prediction $y_t(\beta)$ of the form

$$y_t(\beta) = f_t\left(y_{t-1}, y_{t-2}, ...y_{t-p}, \varepsilon_{t-1}\left(\beta\right), \varepsilon_{t-2}\left(\beta\right), ...\varepsilon_{t-q}\left(\beta\right); \beta\right), \ \ t \in \mathbb{Z},$$

in which β^* is replaced by any β. Assuming $\varepsilon_t(\beta^*) = \varepsilon_t$, it is clear that $y_t(\beta^*) = E\left(y_t/y^{t-1}\right)$ represents the best prediction of y_t (in the mean squares sense) based on data until time $t-1$. Since data are generally available only for $t \geq 1$, the calculation of $\varepsilon_t(\beta)$, for a given β, requires $p+q$ initial values. As the effect of these starting values decrease exponentially for, a causal model, it is possible to approximate $\varepsilon_t(\beta)$ by the following conditional prediction error function, denoted by $\widehat{\varepsilon}_t(\beta)$

$$\begin{aligned} \widehat{\varepsilon}_t(\beta) &= y_t - \widehat{y}_t(\beta), \qquad\qquad t \geq 1 \\ &= y_t - f_t\left(y_{t-1}, y_{t-2}, ...y_{t-p}, \widehat{\varepsilon}_{t-1}\left(\beta\right), \widehat{\varepsilon}_{t-2}\left(\beta\right), ..., \widehat{\varepsilon}_{t-q}\left(\beta\right); \beta\right), \end{aligned}$$

which is conditioned on the initial values $\widehat{\varepsilon}_t(\beta) = y_t = 0$ for $t \leq 0$.

It is worth nothing that any recursive algorithm can be seen as a stochastic analog of a deterministic optimization method. Thus the building of any recursive method starts from an objective function to realize, passes by a choice of a structure of realization and procedures of stochastic optimization, and leads to a precise algorithm for effective implementation on the data. Several criteria can then be interesting to optimize, such as the maximum likelihood criterion, the quadratic criterion or the Bayesian approach. However, the most frequently selected criterion to optimize the choice of parameters in the recursive estimation literature is the minimization of the mean squares error function, i.e. one has to choose a value $\overline{\beta} \in \mathcal{M}$ that minimize

$$V(\beta) = \overline{E}\left(\frac{\widehat{\varepsilon}_t^2(\beta)}{\alpha_t}\right) \tag{3.2.2}$$

where $\mathcal{M}$ is the parameter space, which is generally assumed to be a compact set and (α_t) is a sequences of appropriate weights. The symbol $\overline{E}\,(.)$ denotes the "statistical mean" which is defined by

$$\overline{E}\,(u_t) = \lim_{t\to\infty} \frac{1}{t} \sum_{k=1}^{t} E\,(u_t) \tag{3.2.3}$$

with an implied assumption that the limit exists when the symbol is used. For example, when the sequence $\{\widehat{\varepsilon}_t(\beta)\}$ is second order stationary and (α_t) is stationary (in the sense that $\alpha_t = \alpha$, $\forall t$), the criterion (3.2.2) can be rewritten as follows

$$V(\beta) = \frac{1}{\alpha} E\left(\widehat{\varepsilon}_t^2(\beta)\right)$$

and for a periodically correlated sequence $\{\widehat{\varepsilon}_t(\beta)\}$ with period S, and S-periodic coefficients (α_t), it becomes

$$V(\beta) = \frac{1}{S} \sum_{i=1}^{S} E\left(\frac{\widehat{\varepsilon}_{i+S\tau}^2(\beta)}{\alpha_i}\right)$$

We note in passing that, for the optimality of criterion (3.2.2), the weights sequence (α_t) is generally taken to be the inverse of the innovation variances.

For the latter criterion function, it can be easily shown that the global minimum of $V(\beta)$, $\overline{\beta}$, when it is unique, coincides with the true value of the parameter β^*. However, it may happen that $V\,(\beta)$ does not have a unique global minimum. In that case, we define the set of minimizing arguments as

$$\overline{\mathcal{M}} = \arg\min_{\beta\in\mathcal{M}} V(\beta) = \left\{ \overline{\beta} \in \mathcal{M} / \ V\left(\overline{\beta}\right) = \min_{\beta\in\mathcal{M}} V(\beta) \right\}$$

Thus, the minimization of the criterion function (3.2.2) can lead to a local minimum of $V(\beta)$, thereby causing difficulties.

The theoretical criterion (3.2.2) defines the goal to reach by the most recursive estimation algorithms. However, this criterion can be the subject of a (deterministic) numerical optimization algorithm, provided that one can obtain the function $V(\beta)$ and its differential properties (gradient and Hessian). This requires prior knowledge about the second order statistical properties of the treated processes, which generally depend on the unknown parameter β^* to be estimated and the assumed structure of $\{\varepsilon_t\}$. At this time the identification of the parameter will not be done directly from data and the estimation problem will have no sense. In order to estimate the model parameter directly from data (we are supposed to have t observations) we thus replace the function $V(\beta)$ by a calculable criterion using the available data. A "natural" way to do this is to replace the theoretical criterion in (3.2.2) by its empirical estimate. Thus we calculate the estimate $\widehat{\beta}_t$ that minimize the following weighted least squares function

$$V_t(\beta) = \frac{1}{t}\sum_{k=1}^{t} \frac{\widehat{\varepsilon}_k^2(\beta)}{\alpha_k}, \tag{3.2.4}$$

provided that we can estimate the coefficients (α_t). This approach allows to build, for the unknown parameter β, a sequence of estimates $\widehat{\beta}_t$, making it possible, under some suitable assumptions, to approach more and more the value $\overline{\beta}$ that minimizes (3.2.2). One of the main goals of recursive estimation is to study the behavior of $\widehat{\beta}_t$ when t goes to infinity, and in particular the convergence to the solution $\overline{\beta}$ (or into the set $\overline{\mathcal{M}}$ in the sense that $\inf_{\overline{\beta}\in\overline{\mathcal{M}}} \left|\widehat{\beta}_t - \overline{\beta}\right| \overset{t\to\infty}{\to} 0$ almost surely) and hence to β^*. Certain desirable properties, such as the "compactness" of $\mathcal{M}$, the differentiability of V and the ergodicity of the underlying prediction error process, allow us to expect such a result whenever $V_t(\beta)$ converges (uniformly on β in the sense that $\sup_{\beta\in\mathcal{M}} |V_t(\beta) - V(\beta)| \to 0$ $a.s.$ as $t\to\infty$) to $V(\beta)$ as $t\to\infty$.

In what follows we shall use the weighted least squares criterion (3.2.4) to build our recursive estimation algorithms for $PARMA$ models. First, we begin by estimating the parameters of PAR

models. A generalization to more general $PARMA$ case is then given. The following section is devoted to the adaptation of the best-known and the most widely used algorithm, the recursive least squares algorithm (RLS) to the PAR case.

3.3 Recursive least squares estimation algorithm for PAR models

Consider the periodically correlated process $\{y_t;\ t \in \mathbb{Z}\}$ that satisfies the following $PAR_S(p)$ representation

$$y_t - \phi_{t,1} y_{t-1} - ... - \phi_{t,p} y_{t-p} = \varepsilon_t,\ t \in \mathbb{Z},$$

where $\{\varepsilon_t;\ t \in \mathbb{Z}\}$ is an independent process with zero mean and variance σ_t^2. As mentioned in Section 2.4, the order p can also be considered periodic in time, say p_t, in order to reduce the number of parameters to be estimated. However, in this case a constrained estimation procedure, which is not the objective of this paper, would be needed. Also, we note that the independence assumption on $\{\varepsilon_t\}$ can be replaced by a less stringent assumption which requires this sequence to be only a Martingale difference with periodic variance $E\left(\varepsilon_t^2/\mathcal{F}_{t-1}\right) = \sigma_t^2$ where $\mathcal{F}_t$ denotes the $\sigma-$algebra generated by $\varepsilon_1, \varepsilon_2, ..., \varepsilon_t$. With the above definitions, the PAR model can be written in the form

$$\begin{array}{l} y_{s+S\tau} - \sum\limits_{j=1}^{p} \phi_{s,j}\ y_{s-j+S\tau} = \varepsilon_{s+S\tau}, \\ \text{with } t = s + S\tau,\ s = 1, ..., S \text{ and } \tau = \left\lceil \frac{t-1}{S} \right\rceil \in \mathbb{Z}. \end{array} \tag{3.3.1}$$

Let $\underline{\varphi} = (\underline{\varphi}_1', \underline{\varphi}_2', ..., \underline{\varphi}_S')'$ be the $pS \times 1$ vector of unknown autoregressive parameters, where $\underline{\varphi}_s = (\varphi_{s,1}, \varphi_{s,2}, ..., \varphi_{s,p})'$, $s = 1, 2, ..., S$ and $\underline{\sigma}^2 = (\sigma_1^2, \sigma_2^2, ..., \sigma_S^2)'$ the $S \times 1$ vector of unknown variances of the innovation process $\{\varepsilon_t\}$. Furthermore, we consider the following hypotheses

$H1$: the ergodic process $\{y_t;\ t \in \mathbb{Z}\}$ is such that $E\left(y_i^2 y_j^2\right) < \infty$, for all $i, j \in \mathbb{Z}$,

$H2$: the periodic model (3.3.1) is causal.

Model causality can be established from two existing approaches (see Troutman, 1979; Tiao and Grupe, 1980; Bentarzi and Hallin, 1994; Bentarzi, 1998; Ula and Smadi, 1997, and many others). We note that $H1$ is a univariate version of assumption $A3$ given in Section 2.3.

3.3.1 The $PRLS$ algorithm

Let $y_1, y_2, ..., y_t$ denote the available observations, until time t, of $\{y_t,\ t \in \mathbb{Z}\}$, which satisfies the stochastic difference equation (3.3.1). It is important to establish an estimation procedure to estimate the parameter vectors $\underline{\varphi}$ and $\underline{\sigma}^2$ of this model, that keeps the spatial and temporal complexity invariant, in the sense of (3.2.1), with respect to an increasing t. The adopted optimality criterion in the RLS estimation is, as in most recursive estimation procedures, the mean square error. Thus the problem consists of finding the argument which minimizes

$$V_t\left(\underline{\phi}\right) = \frac{1}{t}\sum_{k=1}^{t}\frac{1}{\sigma_k^2}\widehat{\varepsilon}_k^2\left(\underline{\phi}\right), \tag{3.3.2}$$

where $\widehat{\varepsilon}_t\left(\underline{\varphi}\right) = y_t - \widehat{y}_t\left(\underline{\varphi}\right)$ denotes the error of the one-step-ahead forecast $\widehat{y}_t\left(\underline{\varphi}\right)$ of the future value y_t, based on observed series from 1 to $t-1$. Expression (3.3.2) is the same (except for a multiplicative constant) as the conditional likelihood when $\{\varepsilon_t\}$ is Gaussian. Let $\underline{\mathcal{Y}}_{s+S\tau}$ be the $pS \times 1$ vector of series values given by

$$\underline{\mathcal{Y}}_{s+S\tau} = \left(\underbrace{\mathbf{0}', ..., \mathbf{0}'}_{s-1 \text{ time}}, y_{s+S\tau-1}, ..., y_{s+S\tau-p}, \underbrace{\mathbf{0}', ..., \mathbf{0}'}_{S-s \text{ time}}\right)', \tag{3.3.3}$$

where $\mathbf{0}$ is the $p\times 1$ null vector and where s and τ are such that $t = s+\tau S$, $(1 \leq s \leq S,\ \tau \in \mathbb{Z})$. When $s = 1$ or $s = S$ in (3.3.3), the notation "0 time" means that there are no zero entries in the vector. The start-up values in (3.3.1) are represented by the vector $\underline{y}'_{1,0}$. Using these notations, we can state the following result, which establishes a periodic recursive least squares ($PRLS$) algorithm. This algorithm provides optimal estimators, in the sense of the aforementioned criterion, of the periodic PAR parameters.

Proposition 3.3.1 (Bentarzi and Aknouche, 2006)

The following algorithm, defined via the recursive equations

$$\left\{\begin{array}{l} \widehat{\varepsilon}_{s+S\tau} = y_{s+S\tau} - \underline{\widehat{\phi}}'_{s+S\tau-1}\underline{\mathcal{Y}}_{s+S\tau}, \\ \widehat{\sigma}_j^{2(s,\tau)} = \left\{\begin{array}{l} \dfrac{\tau-\delta}{\tau+1-\delta}\,\widehat{\sigma}_j^{2(s,\tau-1)} + \dfrac{1}{\tau+1-\delta}\,\widehat{\varepsilon}^2_{j+S\tau},\ j \leq s, \\ \dfrac{\tau-1-\delta}{\tau-\delta}\,\widehat{\sigma}_j^{2(s,\tau-1)} + \dfrac{1}{\tau-\delta}\,\widehat{\varepsilon}^2_{j+S(\tau-1)},\ j > s, \end{array}\right. \quad \tau > \delta, \\ \underline{\widehat{\phi}}_{s+S\tau} = \underline{\widehat{\phi}}_{s+S\tau-1} + \dfrac{P_{s+S\tau-1}\underline{\mathcal{Y}}_{s+S\tau}\widehat{\varepsilon}_{s+S\tau}}{\widehat{\sigma}_s^{2(s,\tau)} + \underline{\mathcal{Y}}'_{s+S\tau}P_{s+S\tau-1}\underline{\mathcal{Y}}_{s+S\tau}}, \qquad \underline{\widehat{\phi}}_0 = 0, \\ P_{s+S\tau} = P_{s+S\tau-1} - \dfrac{P_{s+S\tau-1}\underline{\mathcal{Y}}_{s+S\tau}\underline{\mathcal{Y}}'_{s+S\tau}P_{s+S\tau-1}}{\widehat{\sigma}_s^{2(s,\tau)} + \underline{\mathcal{Y}}'_{s+S\tau}P_{s+S\tau-1}\underline{\mathcal{Y}}_{s+S\tau}};\ P_0 = K.I\ ,1 \leq s \leq S,\ \tau \in \mathbb{Z}, \end{array}\right. \tag{3.3.4}$$

provides estimators which minimize (3.3.2), where I denotes the identity matrix. *Here*

$$\widehat{\sigma}_j^{2(s,0)} = \left\{\begin{array}{ll} \widehat{\varepsilon}_j^2 & \textit{if } 1 \leq j \leq s, \\ 0 & \textit{if } \ j > s, \end{array}\right. \quad , \ \delta = (p+1)S,$$

and K is required, for technical purposes in order to obtain an on-line estimator which has the same finite properties as its analogous off-line one, to be a large number.

Proof

The periodic autoregressive model (3.3.1) can be written in the regression form

$$y_{s+S\tau} = \underline{\phi}'\,\underline{\mathcal{Y}}_{s+S\tau} + \varepsilon_{s+S\tau} \tag{3.3.5}$$

It is well known that the best one-step-ahead linear prediction $\widehat{y}_{s+S\tau}\left(\underline{\phi}\right)$ of $y_{s+S\tau}$, based on the past values $y_1, ..., y_{s+S\tau-1}$, is

$$\widehat{y}_{s+S\tau}\left(\underline{\phi}\right) = \underline{\phi}'\,\underline{\mathcal{Y}}_{s+S\tau},$$

from which the objective function can be written as

$$V_{s+S\tau}\left(\underline{\phi}\right) = \frac{1}{s+S\tau}\sum_{k=1}^{s+S\tau}\frac{1}{\sigma_k^2}\left(y_k - \underline{\phi}'\underline{\mathcal{Y}}_k\right)^2.$$

If the variances σ_s^2, $1 \leq s \leq S$, are known, the least squares estimator $\underline{\widehat{\varphi}}_{s+S\tau}$ of $\underline{\varphi}$ is then given, conditional on the initial values $\underline{y}'_{1,0}$, by the normal equation

$$\left(\sum_{k=1}^{s+S\tau}\frac{1}{\sigma_k^2}\,\underline{\mathcal{Y}}_k\underline{\mathcal{Y}}'_k\right)\underline{\widehat{\phi}}_{s+S\tau} = \left(\sum_{k=1}^{s+S\tau}\frac{1}{\sigma_k^2}\,\underline{\mathcal{Y}}_k y_k\right). \tag{3.3.6}$$

The main idea of the desired algorithm is to resolve recursively the normal equation (3.3.6). For this, let

$$R_{s+S\tau} = \sum_{k=1}^{s+S\tau} \frac{1}{\sigma_k^2} \underline{\mathcal{Y}}_k \underline{\mathcal{Y}}_k'.$$

Then

$$R_{s+S\tau} = R_{s+S\tau-1} + \frac{1}{\sigma_s^2} \underline{\mathcal{Y}}_{s+S\tau} \underline{\mathcal{Y}}_{s+S\tau}'. \tag{3.3.7}$$

Using (3.3.6) and (3.3.7), one can, after some simple manipulations, obtain the recursive equation

$$\underline{\widehat{\phi}}_{s+S\tau} = \underline{\widehat{\phi}}_{s+S\tau-1} + \frac{1}{\sigma_s^2} R_{s+S\tau}^{-1} \underline{\mathcal{Y}}_{s+S\tau} \widehat{\varepsilon}_{s+S\tau}, \tag{3.3.8}$$

where

$$\widehat{\varepsilon}_{s+S\tau} = y_{s+S\tau} - \underline{\widehat{\phi}}_{s+S\tau-1}' \underline{\mathcal{Y}}_{s+S\tau}. \tag{3.3.9}$$

Equation (3.3.8) needs, at each iteration, the inversion of a $pS \times pS$ matrix. This is practically undesirable from the point of view of algorithmic complexity which is of order $O\left(p^3 S^3\right)$. To overcome this difficulty, let $P_{s+S\tau} = R_{s+S\tau}^{-1}$ and apply in (3.3.7) and (3.3.8) the Woodbury matrix inversion lemma given by (1.4.13) (see for example, Plackett, 1950; Solo, 1979, p. 961 or Ljung and Söderström, 1983, Lemma 2.1)with $A = P_{s+S\tau-1}$, $B = \underline{\mathcal{Y}}_{s+S\tau}$, $C = \dfrac{1}{\sigma_s^2}$ and $D = \underline{\mathcal{Y}}_{s+S\tau}'$, we obtain (3.3.4).

Recall that the variances σ_s^2, $s = 1, ..., S$, in the above recursions are, until now, assumed to be known. These parameters are typically unknown in practice and need to be estimated. The following recursive formula provides an estimator $\widehat{\sigma}_j^{2(s,\tau)}$ of σ_j^2, for $j = 1, 2, ..., S$, by using the available prediction errors to $t = s + S\tau$. From the well-known least squares theory, a consistent estimator of σ_j^2 (cf, Chapter 1, Section 1.6, or Basawa and Lund, 2001) is given by

$$\begin{aligned}
\widehat{\sigma}_j^{2(s,\tau)} &= \begin{cases} \dfrac{1}{\tau+1-\delta} \sum\limits_{k=0}^{\tau} \widehat{\varepsilon}_{j+Sk}^2 \ , \ j \leq s, \ \tau \geq \delta, \\ \dfrac{1}{\tau-\delta} \sum\limits_{k=0}^{\tau-1} \widehat{\varepsilon}_{j+Sk}^2 \ , \ j > s, \ \tau \geq \delta+1 \end{cases} \\
&= \begin{cases} \dfrac{1}{\tau+1-\delta} \left((\tau-\delta) \widehat{\sigma}_j^{2(s,\tau-1)} + \widehat{\varepsilon}_{j+S\tau}^2 \right), \ j \leq s, \ \tau \geq \delta, \\ \dfrac{1}{\tau-\delta} \left((\tau-\delta-1)\, \widehat{\sigma}_j^{2(s,\tau-1)} + \widehat{\varepsilon}_{j+S\tau}^2 \right), \ j > s, \ \tau \geq \delta+1 \end{cases},
\end{aligned} \tag{3.3.10}$$

with

$$\widehat{\sigma}_j^{2(s,0)} = \begin{cases} \widehat{\varepsilon}_j^2 & j \leq s, \\ 0 & j > s, \end{cases}$$

and $\widehat{\sigma}_j^{2(s,1)} = \widehat{\varepsilon}_j^2$ if $j > s$, where the denominator $(\tau+1-\delta)$ is introduced to obtain an approximately unbiased estimator. In the start-up case where $1 \leq \tau \leq \delta$, we use

$$\widehat{\sigma}_j^{2(s,\tau)} = \begin{cases} \dfrac{\tau}{\tau+1}\widehat{\sigma}_j^{2(s,\tau-1)} + \dfrac{1}{\tau+1}\widehat{\varepsilon}_{j+S\tau}^2, & j \leq s, \\ \dfrac{\tau-1}{\tau}\widehat{\sigma}_j^{2(s,\tau-1)} + \dfrac{1}{\tau}\widehat{\varepsilon}_{j+S(\tau-1)}^2, & j > s, \end{cases} \quad 1 \leq \tau \leq \delta,\ j = 1, ..., S;$$

hence, our work is complete. ■

It is well known in Bayesian time series analysis (e.g. Ljung and Söderström 1983, Lemma 2.2) that the posterior distribution of $\underline{\varphi}$ given a time series until time t, relative to a Gaussian prior distribution with mean $\underline{\varphi}^{(0)}$ and covariance matrix $P^{(0)}$, is also Gaussian. The mean and the covariance matrix of this posterior distribution are, respectively, $\widehat{\underline{\varphi}}_t$ and R_t^{-1} as given by the recursions (3.3.7) and (3.3.8), with initial values $\widehat{\underline{\varphi}}_0 = \underline{\varphi}^{(0)}$ and $R_0^{-1} = P^{(0)}$. Thus, these starting values can be viewed as prior knowledge about $\underline{\varphi}$. This permits one to consider the inverse variances, σ_s^{-2}, $s = 1, 2, ..., S$, of the periodic white noise process $\{\varepsilon_k\}$ as optimal weighting factors in (3.3.2).

It is worth noting that the proposed $PRLS$ algorithm (3.3.4) needs to execute $(p \times S)^2$ multiplications and to memorize a pS dimensional vector and a $pS \times pS$ dimensional matrix. Thus, the algorithmic complexity is of order $O\left((p \times S)^2\right)$.

3.3.2 Asymptotic properties of the $PRLS$ estimator

Concerning the properties of the estimator obtained by the $PRLS$ algorithm, we state the following theorem, which shows under the above conditions that the $PRLS$ estimator is strongly consistent and asymptotically efficient.

Theorem 3.3.1 (Bentarzi and Aknouche, 2006)

i) Under the above assumptions, the $PRLS$ algorithm provides a recursive estimator $\widehat{\underline{\varphi}}_t$ which converges almost surely to the true value $\underline{\varphi}^*$ of the unknown parameter vector $\underline{\varphi}$.

ii) Furthermore, we have

$$\sqrt{s+S\tau}(\widehat{\underline{\phi}}_{s+S\tau} - \underline{\phi}^*) \underset{\tau \longrightarrow \infty}{\overset{D}{\longrightarrow}} N(0, SF^{-1}), \tag{3.3.11}$$

where $F = \sum_{s=1}^{S} \frac{1}{\sigma_s^2} E\left(\underline{\mathcal{Y}}_{s+S\tau}\underline{\mathcal{Y}}'_{s+S\tau}\right)$ *and '*$\overset{D}{\longrightarrow}$*' denotes convergence in distribution.*

Proof

We only prove part ii) of the theorem as the first part is easily obtained by similar techniques. From (3.3.7) and (3.3.8), we can easily verify, by backward recursion, that $\widehat{\underline{\varphi}}_{s+S\tau}$ satisfies the relation

$$\left(R_0 + \sum_{k=1}^{s+S\tau} \frac{1}{\widehat{\sigma}_{s_k}^{2(s_k,\tau_k)}} \underline{\mathcal{Y}}_k \underline{\mathcal{Y}}'_k\right) \widehat{\underline{\phi}}_{s+S\tau} = \left(R_0 \widehat{\underline{\phi}}_0 + \sum_{k=1}^{s+S\tau} \frac{1}{\widehat{\sigma}_{s_k}^{2(s_k,\tau_k)}} \underline{\mathcal{Y}}_k y_k\right), \tag{3.3.12}$$

which is asymptotically equivalent to

$$\left(\sum_{k=1}^{s+S\tau} \frac{1}{\widehat{\sigma}_{s_k}^{2(s_k,\tau_k)}} \underline{\mathcal{Y}}_k \underline{\mathcal{Y}}'_k\right) \widehat{\underline{\phi}}_{s+S\tau} = \sum_{k=1}^{s+S\tau} \frac{1}{\widehat{\sigma}_{s_k}^{2(s_k,\tau_k)}} \underline{\mathcal{Y}}_k y_k. \tag{3.3.13}$$

Note that in (3.3.12), s_k and τ_k are such that $k = s_k + S\tau_k$, $s_k = 1, ..., S$, $\tau_k = \left\lfloor \frac{k-1}{S} \right\rfloor$. Hence, to analyze the asymptotic properties of $\widehat{\underline{\varphi}}_{s+S\tau}$ in (3.3.12), we can use (3.3.13). This yields, with

$$y_{s+S\tau} = \underline{\mathcal{Y}}'_{s+S\tau}\underline{\varphi}^* + \varepsilon_{s+S\tau},$$

the following expression

$$\sqrt{s+S\tau}\left(\widehat{\underline{\phi}}_{s+S\tau} - \underline{\phi}^*\right) = \left(\frac{1}{s+S\tau}\sum_{k=1}^{s+S\tau} \frac{1}{\widehat{\sigma}_{s_k}^{2(s_k,\tau_k)}} \underline{\mathcal{Y}}_k \underline{\mathcal{Y}}'_k\right)^{-1} \left(\frac{1}{\sqrt{s+S\tau}}\sum_{k=1}^{s+S\tau} \frac{1}{\widehat{\sigma}_{s_k}^{2(s_k,\tau_k)}} \underline{\mathcal{Y}}_k \varepsilon_k\right). \tag{3.3.14}$$

To establish (3.3.11), it is enough to show that

$$(s+S\tau)^{-1}\sum_{k=1}^{s+S\tau}\left(\frac{1}{\widehat{\sigma}_{s_k}^{2(s_k,\tau_k)}} - \frac{1}{\sigma_{s_k}^2}\right)\underline{\mathcal{Y}}_k \underline{\mathcal{Y}}'_k = o_p(1), \tag{3.3.15}$$

$$(s+S\tau)^{-1/2}\sum_{k=1}^{s+S\tau}\left(\frac{1}{\widehat{\sigma}_{s_k}^{2(s_k,\tau_k)}} - \frac{1}{\sigma_{s_k}^2}\right)\underline{\mathcal{Y}}_k \varepsilon_k = o_p(1), \tag{3.3.16}$$

from which we deduce that

$$\sqrt{s+S\tau}\left(\widehat{\underline{\phi}}_{s+S\tau}-\underline{\phi}^*\right)=\left(\frac{1}{s+S\tau}\sum_{k=1}^{s+S\tau}\frac{1}{\sigma_k^2}\underline{\mathcal{Y}}_k\underline{\mathcal{Y}}_k'\right)^{-1}\left(\frac{1}{\sqrt{s+S\tau}}\sum_{k=1}^{s+S\tau}\frac{1}{\sigma_k^2}\underline{\mathcal{Y}}_k\varepsilon_k\right)+o_p(1). \quad (3.3.17)$$

Applying the Martingale central limit theorem (cf. Hall and Heyde, 1980, Corollary 3.1) to (3.3.17), (under the independence or the Martingale difference assumption of the innovations process) we find (3.3.11).

First, note that from to the univariate version of Theorem 1.6.2 (see also Theorem 4.2 of Basawa and Lund, 2001)

$$\sqrt{k}(\widehat{\sigma}_{s_k}^{2(s_k,\tau_k)}-\sigma_{s_k}^2)=O_p(1). \quad (3.3.18)$$

Now, to establish (3.3.15), we have

$$(s+S\tau)^{-1}\sum_{k=1}^{s+S\tau}\left(\frac{1}{\widehat{\sigma}_{s_k}^{2(s_k,\tau_k)}}-\frac{1}{\sigma_{s_k}^2}\right)\underline{\mathcal{Y}}_k\underline{\mathcal{Y}}_k' = -(s+S\tau)^{-1}\sum_{k=1}^{s+S\tau}\frac{\left(\widehat{\sigma}_{s_k}^{2(s_k,\tau_k)}-\sigma_{s_k}^2\right)}{\widehat{\sigma}_{s_k}^{2(s_k,\tau_k)}\sigma_{s_k}^2}\underline{\mathcal{Y}}_k\underline{\mathcal{Y}}_k'$$

$$\stackrel{\text{def}}{=} -D. \quad (3.3.19)$$

From (3.3.18), $\widehat{\sigma}_{s_k}^{2(s_k,\tau_k)}$ and $\widehat{\sigma}_{s_k}^{-2(s_k,\tau_k)}$ are bounded in probability. Let $C_1>0$ and $C_2=\max_{s=1,\ldots,S}\left(\sigma_s^{-2}\right)$ be bounding constants, respectively. Because of (3.3.18), $\widehat{\sigma}_{s_k}^{2(s_k,\tau_k)}$ and then $\dfrac{1}{\widehat{\sigma}_{s_k}^{2(s_k,\tau_k)}}$ are bounded in probability. Let C_1 $(C_1>0)$ and $C_2=\max_{k=1,\ldots,S}\left(\dfrac{1}{\sigma_k^2}\right)$ be their bounding constants, respectively. From (3.3.18), (3.3.19) and the Cauchy-Schwarz inequality, we have

$$\begin{aligned}\|D\| &\leq C_1C_2\,(s+S\tau)^{-1}\sum_{k=1}^{s+S\tau}\left|\widehat{\sigma}_{s_k}^{2(s_k,\tau_k)}-\sigma_{s_k}^2\right|\left\|\underline{\mathcal{Y}}_k\right\|^2\\ &\leq C_1C_2\left[\left((s+S\tau)^{-1}\sum_{k=1}^{s+S\tau}\left|\widehat{\sigma}_{s_k}^{2(s_k,\tau_k)}-\sigma_{s_k}^2\right|^2\right)\left((s+S\tau)^{-1}\sum_{k=1}^{s+S\tau}\left\|\underline{\mathcal{Y}}_k\right\|^4\right)\right]^{1/2}.\end{aligned}$$

where $\|.\|$ stands for the Euclidian norm, that is $\|v\|=\sqrt{v'v}$ for any vector v. Since $\left|\widehat{\sigma}_{s_k}^{2(s_k,\tau_k)}-\sigma_{s_k}^2\right|\stackrel{\tau_k\to\infty}{\longrightarrow}0$ for all $s_k=1,\ldots,S$, a Toeplitz lemma gives

$$(s+S\tau)^{-1}\sum_{k=1}^{s+S\tau}\left|\widehat{\sigma}_{s_k}^{2(s_k,\tau_k)}-\sigma_{s_k}^2\right|^2\stackrel{\tau\to\infty}{\longrightarrow}0.$$

From assumptions $H1$ and $H2$, we have

$$\lim_{\tau \longrightarrow \infty} (s + S\tau)^{-1} \sum_{k=1}^{s+S\tau} \left\| \underline{\mathcal{Y}}_k \right\|^4 < \infty$$

Thus (3.3.15) holds.

To prove (3.3.16), we set for some δ such that $0 < \delta < 1/2$

$$S_{s+S\tau} = (s + S\tau)^{-1/2+\delta} \sum_{k=1}^{s+S\tau} \frac{\left(\widehat{\sigma}_{s_k}^{2(s_k,\tau_k)} - \sigma_{s_k}^2 \right)}{\widehat{\sigma}_{s_k}^{2(s_k,\tau_k)} \sigma_{s_k}^2} \underline{\mathcal{Y}}_k \varepsilon_k.$$

Thus, if we can show that $E \left\| S_{s+S\tau} \right\|^2$ is finite for all $s = 1, ..., S$, then Tchebyshev's inequality proves that

$$(s + S\tau)^{-\delta} \left\| S_{s+S\tau} \right\| = o_p(1),$$

which is equivalent to (3.3.16). Indeed, using $H1$ and $H2$ we have

$$\begin{aligned}
E\left(\left\| S_{s+S\tau} \right\|^2 \right) &\leq (s + S\tau)^{-1+2\delta} \sum_{k=1}^{s+S\tau} \frac{E \left\| \left(\widehat{\sigma}_{s_k}^{2(s_k,\tau_k)} - \sigma_{s_k}^2 \right) \underline{\mathcal{Y}}_k \varepsilon_k \right\|^2}{\widehat{\sigma}_{s_k}^{4(s_k,\tau_k)} \sigma_{s_k}^4} \\
&\leq (s + S\tau)^{-1+2\delta} C_1^2 C_2^2 C_3^2 \sum_{k=1}^{s+S\tau} \left(E \left\| \underline{\mathcal{Y}}_k \right\|^2 E\varepsilon_k^2 \right) k^{-1} \\
&\leq \left(C_1^2 C_2^2 C_3^2 \max_{k=1,\ldots,S} E \left\| \underline{\mathcal{Y}}_k \right\|^2 \max_{k=1,\ldots,S} E\varepsilon_k^2 \right) (s + S\tau)^{-1+2\delta} \log(s + S\tau) \\
&< \infty
\end{aligned}$$

where, by (3.3.18), C_3 is a positive number such that $k^{1/2} \left| \widehat{\sigma}_{s_k}^{2(s_k,\tau_k)} - \sigma_{s_k}^2 \right| \leq C_3$ almost surely. Now (3.3.16) is established and completes our proof. ■

Note that the limiting covariance matrix of our on-line estimator $\widehat{\underline{\varphi}}_{s+S\tau}$ coincides with that of the efficient off-line Yule-Walker's estimator (Pagano, 1978) and with that of the maximum likelihood and hence of the weighted least squares estimators (Basawa and Lund, 2001).

Remark 3.3.1

The proof of Theorem 3.3.1-ii) does not straightforwardly follow from the Pagano's (1978) or Basawa and Lund's (2001) proofs. In fact, the proof of part *ii*) is necessary for a good understanding of the major difference in the consideration of the weighting factors in our on-line procedure and those in Pagano's (1978) or Basawa and Lund's (2001). More precisely, the ML estimator of PAR models (which is statistically equivalent to the WLS estimator (see section 16, or Basawa and Lund, 2001) or the Yule-Walker estimator (cf, Pagano, 1978)) can be considered as based on the following normal equation (see Section 1.6, formulae (16.13) or (1.6.19))

$$\left(\sum_{k=1}^{t} \frac{1}{\widehat{\sigma}_k^2(\widehat{\underline{\varphi}}_t)} \underline{\mathcal{Y}}_k \underline{\mathcal{Y}}_k'\right) \widehat{\underline{\varphi}}_t^{(MV)} = \sum_{k=1}^{t} \frac{1}{\widehat{\sigma}_k^2(\widehat{\underline{\varphi}}_t)} \underline{\mathcal{Y}}_k y_k, \tag{3.3.20}$$

in which the weights $\left(\frac{1}{\widehat{\sigma}_k^2(\widehat{\underline{\varphi}}_t)}\right)_{k=1,t}$ are functions of the same estimator $\widehat{\underline{\varphi}}_t$ obtained by using a time series of beforehand fixed size t.

In our on-line case, the $PRLS$ estimator is however based on the normal equation (3.3.13) which we rewrite for convenience as follows

$$\left(\sum_{k=1}^{t} \frac{1}{\widehat{\sigma}_k^2(\widehat{\underline{\varphi}}_k)} \underline{\mathcal{Y}}_k \underline{\mathcal{Y}}_k'\right) \widehat{\underline{\varphi}}_t^{(PRLS)} = \sum_{k=1}^{t} \frac{1}{\widehat{\sigma}_k^2(\widehat{\underline{\varphi}}_k)} \underline{\mathcal{Y}}_k y_k. \tag{3.3.21}$$

In (3.3.21), the weights $\left(\frac{1}{\widehat{\sigma}_k^2(\widehat{\underline{\varphi}}_k)}\right)_{k=1,t}$ are functions of the successive parameters estimates $\left(\widehat{\underline{\varphi}}_k\right)_{k=1,t}$ obtained from a time series of size k, $k = 1, ..., t$. This is the major difference which motivates us to draw the proof $ii)$ of Theorem 3.3.1. Indeed, one can easily show that the ML estimator $\widehat{\underline{\varphi}}_t^{(MV)}$ based on equation (3.3.20) has the same limit properties as the estimator which one obtains when we replace in (3.3.20) the weights $\left(1/\widehat{\sigma}_k^2(\widehat{\underline{\varphi}}_t)\right)_{k=1,t}$ by $\left(1/\sigma_k^2\right)_{k=1,t}$. i.e. the one based on the following equation

$$\left(\sum_{k=1}^{t} \frac{1}{\sigma_k^2} \underline{\mathcal{Y}}_k \underline{\mathcal{Y}}_k'\right) \widehat{\underline{\varphi}}_t = \sum_{k=1}^{t} \frac{1}{\sigma_k^2} \underline{\mathcal{Y}}_k y_k. \tag{3.3.22}$$

The proof of the (asymptotic) equivalence between (3.3.20) and (3.3.22) is classical in the literature and can be established by proving the following relations

$$\begin{aligned} t^{-1} \sum_{k=1}^{t} \left(\frac{1}{\widehat{\sigma}_k^2(\widehat{\underline{\varphi}}_t)} - \frac{1}{\sigma_k^2}\right) \underline{\mathcal{Y}}_k \underline{\mathcal{Y}}_k' &= o_p(1), \\ t^{-1/2} \sum_{k=1}^{t} \left(\frac{1}{\widehat{\sigma}_k^2(\widehat{\underline{\varphi}}_t)} - \frac{1}{\sigma_k^2}\right) \underline{\mathcal{Y}}_k \varepsilon_k &= o_p(1), \end{aligned}$$

which generally follow from the law of large numbers (LLN).

In our on-line case however, the proof of the equivalence between (3.3.21) and (3.3.22) by means of the classical techniques based on the LLN seems difficult. This is the reason for which we have shown the relations (3.3.15) and (3.3.16).

3.3.3 Discussion

In this section, a new algorithm called ($PRLS$) has been developed for an on-line estimation of periodic autoregressive models. This algorithm extends the usual recursive least squares (RLS) established for stationary models. The convergence and asymptotic efficiency properties of our developed algorithm were established. As we can see in Section 8 below, intensive simulations demonstrate good performance and favorably compare to other methods. The algorithm developed here is recommended for its simplicity for parameter estimation in both on-line and off-line data. It should be noted that a fast (regularized) version of the $PRLS$ algorithm can be easily developed along the lines of the ones available in the time-invariant AR models (see, e.g. Houacine, 1991 and references therein), provided that a (periodic) Chandrasehkar-type recursion for a PAR model is obtained (Aknouche and Hamdi, 2006).

Of course, in the case of $PARMA$ models, the $PRLS$ algorithm cannot be applied to estimate the model parameters since the criterion (3.3.2) will not be quadratic with respect to the parameter vector β and no exact least squares solution like (3.3.6) can be given explicitly or recursively. In this case, two directions can be taken to generalize the $PRLS$. The first one consists of writing the given $PARMA$ model in a pseudo regression form (as in (3.3.5)), in which the regressor contains some unobservable variables, and then estimating both the model parameters and the unobservable components in a same recursive algorithm. This approach, known as recursive pseudo linear regression method (thus called by Solo, 1978), will be explored in the next section. The second direction consists of optimizing the given prediction error criterion using several approximations (among which we mention in particular the Taylor formula). This approach known as "recursive predictor error method" will be studied in Section 5.

3.4 Periodic extended least squares algorithms for PARMA models

In this section, we propose to adapt the well-known extended least squares algorithm (ELS), which belongs to the class of pseudo linear regression methods (see Panuska, 1968; Young, 1974; Ljung, 1977a; Solo, 1979; Ljung and Söderström, 1983; Goodwin and Sin, 1984; Duflo, 1990; Adams and Goodwin, 1995; Ljung, 1999) to deal with $PARMA$ models. As we can see below, the pseudo linear regression approach for $PARMA$ parameter estimation consists of writing the $PARMA$ model (1.2.3) in a linear regression form, with respect to the process $\{y_t, t \in \mathbb{Z}\}$, where some regressors are unobservable; and then estimating, by the least squares method, both these unobservable components and the unknown parameters of the underlying model. This problem was already considered by Adams and Goodwin (1995) for $PARMA$ recursive estimation. However, their method was based on the ordinary least squares (OLS) method which does not provide an asymptotically efficient estimator even under the normality assumption. Thus, we propose in this work to slightly modify the Adams and Goodwin (1995) algorithms by using the weighted least squares (WLS) method instead of the OLS one. This can lead to a more efficient algorithm as we show below.

Consider the following univariate $PARMA_S(p,q)$ model

$$y_t - \sum_{j=1}^{p} \phi_{tj}\, y_{t-j} = \varepsilon_t - \sum_{j=1}^{q} \theta_{tj}\, \varepsilon_{t-j},\ t \in \mathbb{Z}, \tag{3.4.1}$$

where $\{\varepsilon_t, t \in \mathbb{Z}\}$ is a periodic white noise with variance σ_t^2. Let $\beta' = \left(\phi_1',\, \theta_1',\, \phi_2',\, \theta_2',\, ...,\, \phi_S',\, \theta_S'\right)$ denote the $(p+q)S \times 1$ vector of parameters where $\phi_i = (\phi_{i,1}, \phi_{i,2} ..., \phi_{i,p})'$ and $\theta_i = (\theta_{i,1}, \theta_{i,2} ..., \theta_{i,q})'$, $i = 1, 2, ..., S$.

3.4.1 The $PELS$ algorithm

For $t = i + S\tau$, $i = 1, 2, ..., S$ and $\tau \in \mathbb{Z}$, let $\varphi_{i+S\tau}$ denote the $S(p+q) \times 1$ "pseudo regressor" vector defined as follows

$$\varphi_{i+S\tau} = \left(\underbrace{\mathbf{0}', ..., \mathbf{0}'}_{i-1 \text{ time}}, y_{i+S\tau-1}, ..., y_{i+S\tau-p}, -\varepsilon_{i+S\tau-1}, ..., -\varepsilon_{i+S\tau-q}, \underbrace{\mathbf{0}', ..., \mathbf{0}'}_{S-i \text{ time}}\right)', \tag{3.4.2}$$

where $\mathbf{0}$ is the $(p+q)\times 1$ null vector, $i=1,2,...,S$ and $\tau\in\mathbb{Z}$. Using these notations one can rewrite model (3.4.1) in the following pseudo linear regression form

$$y_t = \beta' \varphi_t + \varepsilon_t, . \tag{3.4.3}$$

in which the pseudo-regressor vector φ_t contains unobserved components (innovations). If we suppose, for instance, that these unobservable values are known, model (3.4.3) can then be seen as a linear regression model, for which the off-line least squares solution $\widehat{\beta}(t)$ minimizing (3.3.2) is given by the following normal equation

$$\left(\sum_{k=1}^{t} \frac{1}{\sigma_k^2}\varphi_k\varphi_k'\right)\widehat{\beta}(t) = \left(\sum_{k=1}^{t}\frac{1}{\sigma_k^2}\varphi_k y_k\right), \tag{3.4.4}$$

in which the unknown parameter variances σ_k^2, $k=1,..,S$ will be replaced by their estimates $\widehat{\sigma}_k^{2(k,\tau)}$. Let $R_t = \sum_{k=1}^{t}\frac{1}{\widehat{\sigma}_k^{2(i_k,\tau_k)}}\varphi_k\varphi_k'$ be the "weighted" design matrix of the pseudo-linear regression problem (3.4.3) with respect to the time series $\{y_k,\ k=1,...,t\}$, where i_k and τ_k are such that $k=i_k+S\tau_k$. Then one can apply the recursive least squares algorithm ($PRLS$) given in Section 3.3 (see relation (3.3.4)) to solve (3.4.4) recursively, where $\widehat{\beta}_t$ is the recursive solution. This gives the following recursive equations

$$\begin{cases}
\widehat{\varepsilon}_{i+S\tau} = y_{i+S\tau} - \widehat{\beta}'_{i+S\tau-1}\varphi_{i+S\tau}, \\
\widehat{\beta}_{i+S\tau} = \widehat{\beta}_{i+S\tau-1} + \dfrac{1}{\widehat{\sigma}_i^{2(i,\tau)}} R^{-1}_{i+S\tau}\varphi_{i+S\tau}\widehat{\varepsilon}_{i+S\tau}, \\
R_{i+S\tau} = R_{i+S\tau-1} + \dfrac{1}{\widehat{\sigma}_i^{2(i,\tau)}}\varphi_{i+S\tau}\varphi'_{i+S\tau}, \\
\widehat{\sigma}_i^{2(i,\tau)} = \dfrac{\tau-\delta}{\tau+1-\delta}\,\widehat{\sigma}_i^{2(i,\tau-1)} + \dfrac{1}{\tau+1-\delta}\widehat{\varepsilon}^2_{i+S\tau},\ \tau>\delta,
\end{cases} \tag{3.4.5}$$

with starting up values $R_1 = \frac{1}{\widehat{\sigma}_1^{2(1,0)}}\varphi_1\varphi_1'$ and $\widehat{\beta}_1 = R_1^{-1}\varphi_1\frac{y_1}{\widehat{\sigma}_1^{2(1,0)}}$, provided that R_1 is nonsingular, where $\delta = S\,(p+q)$ is the number of PAR and PMA parameters.

However, as the innovations (ε_t) are unknown, the parameter β cannot be forwardly estimated from algorithm (3.4.5) which cannot be applied as it stands. Thus we have first to estimate the innovations ε_t from (3.4.1) using the current estimate $\widehat{\beta}_t$. Let $\widehat{\epsilon}_t$ denote such an estimate. Replacing in (3.4.1) the true parameter β by its current estimation $\widehat{\beta}_t$ and ε_t by its estimate $\widehat{\epsilon}_t$, one obtain the following expression

$$\widehat{\epsilon}_t = y_t - \sum_{j=1}^{p} \widehat{\phi}_{tj}^{(t)}\, y_{t-j} - \sum_{j=1}^{q} \widehat{\theta}_{tj}^{(t)}\, \widehat{\epsilon}_{t-j}, \tag{3.4.6}$$

in which the initial values are taken to be null. Let $\widehat{\varphi}_t$ be an estimate of the pseudo regressor φ_t obtained when replacing ε_t in (3.4.2) by its estimate $\widehat{\epsilon}_t$

$$\widehat{\varphi}_{i+S\tau} = \left(\underbrace{\mathbf{0}', ..., \mathbf{0}'}_{i-1 \text{ time}}, y_{i+S\tau-1}, ..., y_{i+S\tau-p}, -\widehat{\epsilon}_{i+S\tau-1}, ..., -\widehat{\epsilon}_{i+S\tau-q}, \underbrace{\mathbf{0}', ..., \mathbf{0}'}_{S-i \text{ time}}\right)', \tag{3.4.7}$$

where $\mathbf{0}$ is the $(p+q)\times 1$ null vector, $i = 1, 2, ..., S$ and $\tau \in \mathbb{Z}$. Then equation (3.4.6) can be written

$$\widehat{\epsilon}_t = y_t - \widehat{\beta}_t'\widehat{\varphi}_t. \tag{3.4.8}$$

Replacing now the vector φ_t by $\widehat{\varphi}_t$ in (3.4.5), one obtains the following periodic extended least squares algorithm

$$\begin{cases} \widehat{\varepsilon}_{i+S\tau} = y_{i+S\tau} - \widehat{\beta}'_{i+S\tau-1}\widehat{\varphi}_{i+S\tau}, \\ \widehat{\sigma}_i^{2(i,\tau)} = \dfrac{\tau-\delta}{\tau+1-\delta}\,\widehat{\sigma}_i^{2(i,\tau-1)} + \dfrac{1}{\tau+1-\delta}\widehat{\varepsilon}^2_{i+S\tau}, \ \tau > \delta, \\ \widehat{\beta}_{i+S\tau} = \widehat{\beta}_{i+S\tau-1} + \dfrac{1}{\widehat{\sigma}_i^{2(i,\tau)}}\widehat{R}^{-1}_{i+S\tau}\widehat{\varphi}_{i+S\tau}\widehat{\varepsilon}_{i+S\tau}, \\ \widehat{R}_{i+S\tau} = \widehat{R}_{i+S\tau-1} + \dfrac{1}{\widehat{\sigma}_i^{2(i,\tau)}}\widehat{\varphi}_{i+S\tau}\widehat{\varphi}'_{i+S\tau}, \\ \widehat{\epsilon}_{i+S\tau} = y_{i+S\tau} - \widehat{\beta}'_{i+S\tau}\widehat{\varphi}_{i+S\tau}, \end{cases} \tag{3.4.9}$$

where $\widehat{\varphi}_t$ is computed from (3.4.7), $\widehat{R}_t = \sum_{k=1}^{t} \frac{1}{\widehat{\sigma}_k^{2(i_k,\tau_k)}}\widehat{\varphi}_k\widehat{\varphi}'_k$ and the starting values are given by $\widehat{R}_1 = \frac{1}{\widehat{\sigma}_1^{2(1,0)}}\widehat{\varphi}_1\widehat{\varphi}'_1$ and $\widehat{\beta}_1 = \widehat{R}_1^{-1}\widehat{\varphi}_1\frac{y_1}{\widehat{\sigma}_1^{2(1,0)}}$. Note that it is possible to improve the algorithmic complexity of the last algorithm (3.4.9) by setting $\widehat{P}_t = \widehat{R}_t^{-1}$ and updating $\widehat{P}_t$ using again the matrix inversion lemma (MIL) (cf, Appendix A). Moreover, as for the $PRLS$ algorithm, to avoid the singularity problem in the calculation of the starting up value $\widehat{\beta}_1 = \widehat{R}_1^{-1}\widehat{\varphi}_1\frac{y_1}{\widehat{\sigma}_1^{2(1,0)}}$, and to obtain an estimate closer to the off-line solution of (3.4.4) we use the approximation, $\widehat{\beta}_0 = 0$ and $\widehat{R}_0 = K^{-1}I$, where K is a large enough number. This gives the following algorithm.

Algorithm 3.4.1 (The $PELS$ algorithm)

The periodic extended least squares ($PELS$) *algorithm for* $PARMA$ *models is given by the following recursive relations*

$$\left\{\begin{array}{l} \widehat{\varepsilon}_{i+S\tau} = y_{i+S\tau} - \widehat{\beta}'_{i+S\tau-1}\widehat{\varphi}_{i+S\tau}, \\ \widehat{\sigma}_i^{2(i,\tau)} = \left\{\begin{array}{l} \dfrac{\tau}{\tau+1}\,\widehat{\sigma}_i^{2(i,\tau-1)} + \dfrac{1}{\tau+1}\widehat{\varepsilon}^2_{i+S\tau},\ 0<\tau\leq\delta, \\ \dfrac{\tau-\delta}{\tau+1-\delta}\,\widehat{\sigma}_i^{2(i,\tau-1)} + \dfrac{1}{\tau+1-\delta}\widehat{\varepsilon}^2_{i+S\tau},\ \tau>\delta, \end{array}\right. \\ \widehat{\beta}_{i+S\tau} = \widehat{\beta}_{i+S\tau-1} + \dfrac{\widehat{P}_{i+S\tau}\widehat{\varphi}_{i+S\tau}\widehat{\varepsilon}_{i+S\tau}}{\widehat{\sigma}_i^{2(i,\tau)} + \widehat{\varphi}'_{i+S\tau}\widehat{P}_{i+S\tau-1}\widehat{\varphi}_{i+S\tau}},\ \widehat{\beta}_0 = 0, \\ \widehat{P}_{i+S\tau} = \widehat{P}_{i+S\tau-1} - \dfrac{\widehat{P}_{i+S\tau-1}\widehat{\varphi}_{i+S\tau}\widehat{\varphi}'_{i+S\tau}\widehat{P}_{i+S\tau-1}}{\widehat{\sigma}_i^{2(i,\tau)} + \widehat{\varphi}'_{i+S\tau}\widehat{P}_{i+S\tau-1}\widehat{\varphi}_{i+S\tau}},\ P_0 = K.I, \\ \widehat{\epsilon}_{i+S\tau} = y_{i+S\tau} - \widehat{\beta}'_{i+S\tau}\widehat{\varphi}_{i+S\tau}, \\ \text{compute } \widehat{\varphi}_{i+S\tau} \text{ from (3.4.7)}, \end{array}\right. \tag{3.4.10}$$

where $\delta = (p+q)\,S$*,* K *is a large enough number and* $\widehat{\sigma}_i^{2(i,0)} = \widehat{\varepsilon}_i^2$*.*

Remark 3.4.1

The main difference between Adams and Goodwin (1995)'s algorithm (Algorithm 1, p130) and our proposed $PELS$ one lies in the fact that the first one is based on the ordinary least squares method which is shown to provide (asymptotically) inefficient estimators for the $PARMA$ case, whereas the $PELS$ algorithm is based on the weighted least squares method (see section 1.6 or Basawa and Lund, 2001) which might lead to more efficient estimates. Moreover the pseudo regressor $\widehat{\varphi}_{i+S\tau}$ expression used in the $PELS$ algorithm (see equation (3.4.7)) is slightly different to the one given in the Adams and Goodwin (1995) algorithms (see equations (8) and (9) in Adams and Goodwin (1995)).

Remark 3.4.2

A principle part of the algorithm and proofs given below is the distinction between the estimated predictive error and residual error, that is

$$\begin{aligned} \text{estimated prediction error } \widehat{\varepsilon}_t &= y_t - \widehat{\beta}'_{t-1}\widehat{\varphi}_t, \\ \text{estimated residuals } \widehat{\epsilon}_t &= y_t - \widehat{\beta}'_t\widehat{\varphi}_t. \end{aligned}$$

The difference in the above is the parameter estimates used.

Remark 3.4.3

A major advantage of the $PELS$ algorithm is its simple algorithmic complexity which is comparable to that of the $PRLS$ algorithm.

3.4.2 Consistency of the $PELS$ estimator

To study the convergence property of the estimator obtained from algorithm (3.4.10), one can use one of the well-known approaches, in the on-line estimation literature, such as the Ordinary Differential Equation approach (Ljung 1977a, 1977b), the Hannan (1976, 1980)'s approach or the stochastic Lyapunov function approach due to Moore and Ledwich (1980) and Solo (1979). However, we adopt the last approach, which is proved to assure the strong convergence property of the recursive estimator without being monitored, i.e., to be projected into a stability region when it is not admissible (stable). The mentioned approach introduces a function (of the estimate) that plays the role of a stochastic Lyapunov function for the recursive algorithm (3.4.10) and then uses Martingale theory to study the convergence of this function, which implies the consistency of the estimate.

Consider the following assumptions :

$B1a$: The $PARMA$ model (3.4.1) is causal, invertible and the process $\{y_t, t \in \mathbb{Z}\}$ is periodically ergodic in the sense that

$$\lim_{t\to\infty} \sup \frac{1}{t} \sum_{k=1}^{t} y_k^2 < \infty, \tag{3.4.11a}$$

$B1b$: The $PARMA$ process $\{y_t, t \in \mathbb{Z}\}$ is such that

$$\lim_{t\to\infty} \sup \frac{1}{t} \sum_{k=1}^{t} y_k^4 < \infty, \tag{3.4.11b}$$

$B2$: The polynomials $\theta_s(L) = 1 - \sum_{i=1}^{q} \theta_{s,j} L^j$, are such that

$$\mathrm{Re}\left[\frac{1}{\theta_s(e^{i\omega})} - \frac{1}{2}\right] > 0 \ , \ \ \forall \omega \in \left]-\pi, \pi\right[\ , \text{ for all } s = 1, ..., S. \tag{3.4.12a}$$

$B3$: The limiting matrix $\lim_{t\to\infty} \frac{1}{t} \sum_{k=1}^{t} \frac{1}{\sigma_k^2} \varphi_k \varphi_k' = R$ exists and is positive definite.

Remark 3.4.4

Note that similar assumptions have been considered by Adams and Goodwin (1995). However, instead of the assumption $B2$, they assumed that a certain related model given by (3.4.34) is very strictly passive (cf. Hill and Moylan, 1980, Moore and Ledwich, 1980). Since this assumption is difficult to check, they also gave a frequency domain sufficient condition for this model to be very strictly passive. This property can be used here, but we prefer the condition $B2$ which seems much simpler to check and so weaker than the sufficient condition of Adams and Goodwin (1995). Further discussions about this point will be given in the following subsections.

Remark 3.4.5

Notice that assumption $B2$ is equivalent to

$$\left|\theta_s(e^{i\omega}) - 1\right| < 1 \ \forall\omega \in \left]-\pi, \pi\right[\text{ , for all } s = 1, ..., S, \tag{3.4.12b}$$

which is more easy to check. To see this equivalence we know, on the one hand that

$$\begin{aligned} \left[\operatorname{Re}\left[\frac{1}{z} - \frac{1}{2}\right] > 0\right] \quad &\Leftrightarrow \quad \frac{2\operatorname{Re} z - |z|^2}{|z|^2} > 0, \\ &\Leftrightarrow \quad 2\operatorname{Re} z > |z|^2 , \end{aligned}$$

and on the other hand, that

$$\begin{aligned} [|z-1| < 1] \quad &\Leftrightarrow \quad [(z-1)(\overline{z}-1) < 1] , \\ &\Leftrightarrow \quad |z|^2 - 2\operatorname{Re} z + 1 < 1, \end{aligned}$$

which establish the result (see Ljung 1977a for further discussions).

For example, for a $PARMA_S\,(p,q)$ model with $q = 1$, condition (3.4.12b) means that

$$\sup_{s=1,...,S} |\theta_{s,1}| < 1,$$

which is more stringent than the invertibility condition

$$\prod_{s=1}^{S} \theta_{s,1} < 1.$$

It should be noted that because of the $\mathcal{F}_t-$measurability of the included weights $\widehat{\sigma}_{i_t}^{-2(i_t,\tau_t)}$, the $PELS$ algorithm as it stands, does not guarantee the stability of $\widehat{\theta}_t^{(t-1)}(L)$ and hence we must prevent the recursive estimate $\widehat{\beta}_t$ from getting outside the stability region. This may be accomplished by including a projection facility in the algorithm as follows

$$\widehat{\beta}_{i+S\tau} = \left[\widehat{\beta}_{i+S\tau-1} + \frac{\widehat{P}_{i+S\tau}\widehat{\psi}_{i+S\tau}\widehat{\varepsilon}_{i+S\tau}}{\widehat{\sigma}_i^{2(i,\tau)} + \widehat{\psi}'_{i+S\tau}\widehat{P}_{i+S\tau-1}\widehat{\psi}_{i+S\tau}}\right]_{\mathcal{M}_S},$$

where $\mathcal{M}_S$ denotes such a stability region, i.e. the region of causality and of invertibility, and

$$[\beta]_{\mathcal{M}_S} = \begin{cases} \beta & \text{if } \beta \in \mathcal{M}_S \\ \text{a value strictly interior to } \mathcal{M}_S & \text{otherwise.} \end{cases}$$

Practically, for each value t, we must monitor the admissibility of the values $\widehat{\beta}_t$., i.e. the autoregressive and moving average polynomials must satisfy the causality and invertibility conditions. Ljung and Söderström (1983) proposed various approaches, which consist of testing the admissibility of the current estimate and in the event of rejection to apply a projection of the estimate into $\mathcal{M}_S$, where the values are acceptable. Such a projection is implemented as successive reductions of the correction term for the parameter estimates. We apply the following method to monitor $\widehat{\beta}_t$ (cf, Ljung and Söderström, 1983, p. 367).

Monitoring algorithm

i) Compute $u_t = \dfrac{\widehat{P}_t\widehat{\psi}_t\widehat{\varepsilon}_t}{\widehat{\sigma}_{i_t}^{2(i_t,\tau)} + \widehat{\psi}_t'\widehat{P}_{t-1}\widehat{\psi}_t}$ (the updating term).

ii) Compute $\widehat{\beta}_t = \widehat{\beta}_{t-1} + u_t$.

ii) Test if $\widehat{\beta}_t \in M_S$ *(is admissible). If yes stop.*

iii) If $\widehat{\beta}_t \notin M_S$*, set* $u_t \longleftarrow \lambda u_t$*, for some* $0 \leq \lambda < 1$ *and go to ii).*

The forgoing algorithm is applied until we obtain an admissible value. Further discussion on the choice of λ is given in Ljung and Söderström (1983, chapter 5). It is worth noting that the projection assumption does not imply that the projection is necessary for the practical use of the algorithm. Thus for some variants of the $PELS$ we make the following additional assumption

$B4$: The algorithm includes a projection that keeps $\widehat{\beta}_t$ in the stability region (at least infinitely often).

The following theorem shows the strong consistency of the $PELS$ estimator.

Theorem 3.4.1

$i)$ Under assumptions $B1$, $B2$ and $B4$ the *obtained* $PELS$ *estimator* $\widehat{\beta}_t$ *converges almost surely* *to the true value* β^**, in the sense that*

$$\frac{1}{t}\left(\widehat{\beta}_t - \beta^*\right)' \widehat{R}_t \left(\widehat{\beta}_t - \beta^*\right) \xrightarrow{a.s.} 0 \qquad as\ t \to \infty, \tag{3.4.13a}$$

$$\frac{1}{t}\sum_{k=1}^{t}\left(\widehat{\epsilon}_k - \varepsilon_k\right)^2 \xrightarrow{a.s.} 0 \quad as\ t \to \infty. \tag{3.4.13b}$$

ii) If, in addition, $B3$ *holds then we have*

$$\widehat{\beta}_t \xrightarrow{a.s.} \beta^*. \tag{3.4.14}$$

iii) For the $PELS$ *algorithm* (3.4.10) *with the weights* $\widehat{\sigma}_{i_t}^{-2(i_t,\tau_t-1)}$ *instead of* $\widehat{\sigma}_{i_t}^{2(i_t,\tau_t)}$*, the obtained estimator satisfies the conclusions i) and ii) without the assumptions* $B1b$ *and* $B4$. ■

The proof of this theorem is based on the stochastic Lyapunov function approach (cf, Solo, 1979; Moore and Ledwich, 1980) which lies on a version of the Robbins and Siegmund's (1971) Martingale convergence theorem, slightly modified by Solo (1979). This theorem can also be found in Neveu (1972, p.33) or in Duflo (1990). The proof is quite long and requires the several necessary lemmas given below. We first begin with the basic Robbins and Siegmund's (1971) convergence results, modified by Solo (1979), which we give without proof.

Lemma 3.4.1 (Robbins and Siegmund, 1971; Solo, 1979)

Let $\{L_t\}$, $\{\alpha_{t+1}\}$, $\{\beta_{t+1}\}$, $\{\gamma_{t+1}\}$ *be sequences of nonnegative random variables adapted to a filtration* $\{\mathcal{F}_t\}$*, such that*

$$E(L_t/\mathcal{F}_{t-1}) \leq (1+\gamma_t)\, L_{t-1} + \alpha_t - \beta_t, \tag{3.4.15a}$$

$$\sum_{t=1}^{\infty}\gamma_t < \infty \qquad a.s.,$$

$$\sum_{t=1}^{\infty}\alpha_t < \infty \qquad a.s. \tag{3.4.15b}$$

Then

$$L_t \to L \quad a.s., \tag{3.4.16a}$$

and

$$\sum_{t=1}^{\infty}\beta_t < \infty \quad a.s., \tag{3.4.16b}$$

where L is a nonnegative random variable.

Proof

See Solo (1979, p. 961) for the particular case $\gamma_t = 0$ and Duflo (1990) for a general proof. ■

In order to apply the forgoing result to show the strong consistency of the $PELS$ estimator, we need to introduce some notations. Set

$$\begin{aligned}
&\widetilde{\beta}_t = \widehat{\beta}_t - \beta^*,\\
&T_t = \widetilde{\beta}_t' \widehat{R}_t \widetilde{\beta}_t,\\
&w_{i,\tau}^2 = \widehat{\sigma}_i^{2(i,\tau)},\\
&w_{i,\tau}^2 = \widehat{\sigma}_i^{2(i,\tau)},\\
&a_t = -w_{i_t,\tau_t}^{-1} \widehat{\varphi}_t' \widetilde{\beta}_t
\end{aligned}$$

and

$$b_t = w_{i_t,\tau_t}^{-1} \left(\widehat{\epsilon}_t - \varepsilon_t - \frac{1}{2} w_{i_t,\tau_t} a_t \right).$$

Then one can prove the following lemma.

Lemma 3.4.2

The following inequality

$$T_t \leq T_{t-1} - 2a_t b_t + 2w_{i_t,\tau_t}^{-2} \widehat{\varphi}_t' \widetilde{\beta}_{t-1} \varepsilon_t + 2w_{i_t,\tau_t}^{-2} \widehat{\varphi}_t' \widehat{R}_t^{-1} \widehat{\varphi}_t \widehat{\varepsilon}_t \varepsilon_t, \tag{3.4.17}$$

is true for any t.

Proof

For convenience, we rewrite using the above notations, the first, the third and the fourth equations of the $PELS$ algorithm (3.4.9)

$$\widehat{\varepsilon}_t = y_t - \widehat{\beta}_{t-1}' \widehat{\varphi}_t, \tag{3.4.18a}$$

$$\widehat{\beta}_t = \widehat{\beta}_{t-1} + w_{i_t,\tau_t}^{-2} \widehat{R}_t^{-1} \widehat{\varphi}_t \widehat{\varepsilon}_t, \tag{3.4.18b}$$

$$\widehat{R}_t = \widehat{R}_{t-1} + w_{i_t,\tau_t}^{-2} \widehat{\varphi}_t \widehat{\varphi}_t'. \tag{3.4.18c}$$

These equations can be written, with the fact that $\widetilde{\beta}_t = \widehat{\beta}_t - \beta^*$, as follows

$$\begin{aligned}
\widehat{\varepsilon}_t &= y_t - \widehat{\beta}'_{t-1}\widehat{\varphi}_t, \\
\widetilde{\beta}_t &= \widetilde{\beta}_{t-1} + w_{i_t,\tau_t}^{-2}\widehat{R}_t^{-1}\widehat{\varphi}_t\widehat{\varepsilon}_t, \\
\widehat{R}_t &= \widehat{R}_{t-1} + w_{i_t,\tau_t}^{-2}\widehat{\varphi}_t\widehat{\varphi}'_t.
\end{aligned} \tag{3.4.19}$$

Otherwise, from (3.4.8), (3.4.18a) and (3.4.18b) we have

$$\begin{aligned}
\widehat{\epsilon}_t &= y_t - \widehat{\varphi}'_t\widehat{\beta}_t, \\
&= y_t - \widehat{\varphi}'_t\widehat{\beta}_{t-1} - w_{i_t,\tau_t}^{-2}\widehat{\varphi}'_t\widehat{R}_t^{-1}\widehat{\varphi}_t\widehat{\varepsilon}_t, \\
&= \left(1 - w_{i_t,\tau_t}^{-2}\widehat{\varphi}'_t\widehat{R}_t^{-1}\widehat{\varphi}_t\right)\widehat{\varepsilon}_t,
\end{aligned} \tag{3.4.20}$$

and by the matrix inversion lemma (see Appendix A, equation ($A3$))

$$\frac{\widehat{R}_t^{-1}\widehat{\varphi}_t}{w_{i_t,\tau_t}^2 - \widehat{\varphi}'_t\widehat{R}_t^{-1}\widehat{\varphi}_t} = w_{i_t,\tau_t}^{-2}\widehat{R}_{t-1}^{-1}\widehat{\varphi}_t. \tag{3.4.21}$$

Hence, using (3.4.20) and (3.4.21), equation (3.4.19) becomes

$$\begin{aligned}
\widetilde{\beta}_t &= \widetilde{\beta}_{t-1} + \frac{\widehat{R}_t^{-1}\widehat{\varphi}_t\widehat{\epsilon}_t}{w_{i_t,\tau_t}^2 - \widehat{\varphi}'_t\widehat{R}_t^{-1}\widehat{\varphi}_t}, \\
&= \widetilde{\beta}_{t-1} + w_{i_t,\tau_t}^{-2}\widehat{R}_{t-1}^{-1}\widehat{\varphi}_t\widehat{\epsilon}_t,
\end{aligned}$$

this permits us to rewrite (3.4.18b) and (3.4.18c)

$$\widetilde{\beta}_t = \widetilde{\beta}_{t-1} + w_{i_t,\tau_t}^{-2}\widehat{R}_{t-1}^{-1}\widehat{\varphi}_t\widehat{\epsilon}_t, \tag{3.4.22a}$$

$$\widehat{R}_t = \widehat{R}_{t-1} + w_{i_t,\tau_t}^{-2}\widehat{\varphi}_t\widehat{\varphi}'_t. \tag{3.4.22b}$$

Pre-multiplying now (3.4.22a) by $\widehat{R}_{t-1}$ one find

$$\widehat{R}_{t-1}\widetilde{\beta}_t = \widehat{R}_{t-1}\widetilde{\beta}_{t-1} + w_{i_t,\tau_t}^{-2}\widehat{\varphi}_t\widehat{\epsilon}_t,$$

and using the relation (3.4.22b) in the left member of this last equation, it gives

$$\widehat{R}_t\widetilde{\beta}_t - w_{i_t,\tau_t}^{-2}\widehat{\varphi}_t\widehat{\varphi}'_t\widetilde{\beta}_t = \widehat{R}_{t-1}\widetilde{\beta}_{t-1} + w_{i_t,\tau_t}^{-2}\widehat{\varphi}_t\widehat{\epsilon}_t. \tag{3.4.23}$$

Pre-multiplying again (3.4.23) by $\widetilde{\beta}'_t$ we obtain

$$\widetilde{\beta}'_t\widehat{R}_t\widetilde{\beta}_t = \widetilde{\beta}'_t\widehat{R}_{t-1}\widetilde{\beta}_{t-1} + w^{-2}_{i_t,\tau_t}\widetilde{\beta}'_t\widehat{\varphi}_t\widehat{\epsilon}_t + w^{-2}_{i_t,\tau_t}\left(\widetilde{\beta}'_t\widehat{\varphi}_t\right)^2, \tag{3.4.24}$$

and using (3.4.22a) in the first term of the right member of (3.4.24) it follows

$$\widetilde{\beta}'_t\widehat{R}_t\widetilde{\beta}_t = \widetilde{\beta}'_{t-1}\widehat{R}_{t-1}\widetilde{\beta}_{t-1} + w^{-2}_{i_t,\tau_t}\widehat{\varphi}'_t\widetilde{\beta}_{t-1}\widehat{\epsilon}_t + w^{-2}_{i_t,\tau_t}\widetilde{\beta}'_t\widehat{\varphi}_t\widehat{\epsilon}_t + w^{-2}_{i_t,\tau_t}\left(\widetilde{\beta}'_t\widehat{\varphi}_t\right)^2. \tag{3.4.25}$$

Substituting the expression of $\widetilde{\beta}_{t-1}$ from (3.4.22a) into the second term of the right member of (3.4.25), with the notation

$$\begin{aligned} T_t &= \widetilde{\beta}'_t\widehat{R}_t\widetilde{\beta}_t, \\ a_t &= -w^{-1}_{i_t,\tau_t}\widehat{\varphi}'_t\widetilde{\beta}_t, && (3.4.26a) \\ b_t &= w^{-1}_{i_t,\tau_t}\left(\widehat{\epsilon}_t - \varepsilon_t - \frac{1}{2}w_{i_t,\tau_t}a_t\right), && (3.4.26b) \end{aligned}$$

one obtains

$$\begin{aligned} T_t &= T_{t-1} + w^{-2}_{i_t,\tau_t}\widehat{\varphi}'_t\widetilde{\beta}_t\widehat{\epsilon}_t - w^{-4}_{i_t,\tau_t}\widehat{\varphi}'_t\widehat{R}^{-1}_{t-1}\widehat{\varphi}_t\widehat{\epsilon}^2_t + w^{-2}_{i_t,\tau_t}\widetilde{\beta}'_t\widehat{\varphi}_t\widehat{\epsilon}_t + w^{-2}_{i_t,\tau_t}\left(\widetilde{\beta}'_t\widehat{\varphi}_t\right)^2, \\ &= T_{t-1} + 2w^{-2}_{i_t,\tau_t}\widehat{\varphi}'_t\widetilde{\beta}_t\left(\widehat{\epsilon}_t + \frac{1}{2}\widehat{\varphi}'_t\widetilde{\beta}_t\right) - w^{-4}_{i_t,\tau_t}\widehat{\varphi}'_t\widehat{R}^{-1}_{t-1}\widehat{\varphi}_t\widehat{\epsilon}^2_t, \\ &= T_{t-1} + 2w^{-2}_{i_t,\tau_t}\widehat{\varphi}'_t\widetilde{\beta}_t\left(\widehat{\epsilon}_t - \varepsilon_t + \frac{1}{2}\widehat{\varphi}'_t\widetilde{\beta}_t\right) + 2w^{-2}_{i_t,\tau_t}\widehat{\varphi}'_t\widetilde{\beta}_t\varepsilon_t - w^{-4}_{i_t,\tau_t}\widehat{\varphi}'_t\widehat{R}^{-1}_{t-1}\widehat{\varphi}_t\widehat{\epsilon}^2_t, \\ &= T_{t-1} - 2a_tb_t + 2w^{-2}_{i_t,\tau_t}\widehat{\varphi}'_t\widetilde{\beta}_t\varepsilon_t - w^{-4}_{i_t,\tau_t}\widehat{\varphi}'_t\widehat{R}^{-1}_{t-1}\widehat{\varphi}_t\widehat{\epsilon}^2_t. && (3.4.27) \end{aligned}$$

Exploiting (3.4.19) in the third term of the right member of (3.4.27), we find

$$T_t = T_{t-1} - 2a_tb_t + 2w^{-2}_{i_t,\tau_t}\widehat{\varphi}'_t\widetilde{\beta}_{t-1}\varepsilon_t + 2w^{-2}_{i_t,\tau_t}\widehat{\varphi}'_t\widehat{R}^{-1}_t\widehat{\varphi}_t\widehat{\varepsilon}_t\varepsilon_t - w^{-4}_{i_t,\tau_t}\widehat{\varphi}'_t\widehat{R}^{-1}_{t-1}\widehat{\varphi}_t\widehat{\epsilon}^2_t, \tag{3.4.28}$$

and using the fact that the last term of the right member of (3.4.28) is nonnegative, one deduces (3.4.17). ■

Let $\overline{T}_t = \dfrac{1}{t}\left(T_t + 2\sum\limits_{k=1}^{t} a_kb_k\right)$, then the following result can be obtained.

Lemma 3.4.3

Under assumptions B1 and B2 we have

$$E(\overline{T}_t/\mathcal{F}_{t-1}) \leq \overline{T}_{t-1} - t^{-1}\overline{T}_{t-1} + 2t^{-1}c_t^{-1}\sigma_t^2 w_{i_t,\tau_t-1}^{-2}\widehat{\varphi}_t'\widehat{R}_t^{-1}\widehat{\varphi}_t + Ct^{-2}\times \left[|y_t - \widehat{\varepsilon}_t| E\left(\left|\widehat{\varepsilon}_t^2\varepsilon_t\right|/\mathcal{F}_{t-1}\right) + \widehat{\varphi}_t'\widehat{R}_t^{-1}\widehat{\varphi}_t E\left(\left|\widehat{\varepsilon}_t^3\varepsilon_t\right|/\mathcal{F}_{t-1}\right)\right] \quad a.s., \tag{3.4.29}$$

where

$$c_t = \begin{cases} \dfrac{\tau_t}{\tau_t+1} & \text{if } 0 < \tau_t \leq \delta, \\ \dfrac{\tau_t-\delta}{\tau_t+1-\delta} & \text{if } \quad \tau_t > \delta. \end{cases}$$

Proof

Rewrite the inequality (3.4.17) as follows

$$\begin{aligned} T_t &\leq T_{t-1} - 2a_tb_t + 2c_t^{-1}w_{i_t,\tau_t-1}^{-2}\widehat{\varphi}_t'\widetilde{\beta}_{t-1}\varepsilon_t + 2c_t^{-1}w_{i_t,\tau_t-1}^{-2}\widehat{\varphi}_t'\widehat{R}_t^{-1}\widehat{\varphi}_t\widehat{\varepsilon}_t\varepsilon_t \\ &\quad + 2\left(w_{i_t,\tau_t}^{-2} - c_t^{-1}w_{i_t,\tau_t-1}^{-2}\right)\widehat{\varphi}_t'\widetilde{\beta}_{t-1}\varepsilon_t + 2\left(w_{i_t,\tau_t}^{-2} - c_t^{-1}w_{i_t,\tau_t-1}^{-2}\right)\widehat{\varphi}_t'\widehat{R}_t^{-1}\widehat{\varphi}_t\widehat{\varepsilon}_t\varepsilon_t, \\ &= T_{t-1} - 2a_tb_t + 2c_t^{-1}w_{i_t,\tau_t-1}^{-2}\widehat{\varphi}_t'\widetilde{\beta}_{t-1}\varepsilon_t + 2c_t^{-1}w_{i_t,\tau_t-1}^{-2}\widehat{\varphi}_t'\widehat{R}_t^{-1}\widehat{\varphi}_t\left(\widehat{\varepsilon}_t - \varepsilon_t\right)\varepsilon_t \\ &\quad + 2w_{i_t,\tau_t-1}^{-2}\widehat{\varphi}_t'\widehat{R}_t^{-1}\widehat{\varphi}_t\varepsilon_t^2 + 2\left(\frac{c_tw_{i_t,\tau_t-1}^2 - w_{i_t,\tau_t}^2}{c_tw_{i_t,\tau_t}^2w_{i_t,\tau_t-1}^2}\right)\widehat{\varphi}_t'\widetilde{\beta}_{t-1}\varepsilon_t \\ &\quad + 2\left(\frac{c_tw_{i_t,\tau_t-1}^2 - w_{i_t,\tau_t}^2}{c_tw_{i_t,\tau_t}^2w_{i_t,\tau_t-1}^2}\right)\widehat{\varphi}_t'\widehat{R}_t^{-1}\widehat{\varphi}_t\widehat{\varepsilon}_t\varepsilon_t, \\ &\leq T_{t-1} - 2a_tb_t + 2c_t^{-1}w_{i_t,\tau_t-1}^{-2}\widehat{\varphi}_t'\widetilde{\beta}_{t-1}\varepsilon_t + 2c_t^{-1}w_{i_t,\tau_t-1}^{-2}\widehat{\varphi}_t'\widehat{R}_t^{-1}\widehat{\varphi}_t\left(\widehat{\varepsilon}_t - \varepsilon_t\right)\varepsilon_t \\ &\quad + 2c_t^{-1}w_{i_t,\tau_t-1}^{-2}\widehat{\varphi}_t'\widehat{R}_t^{-1}\widehat{\varphi}_t\varepsilon_t^2 + 2\left|\left(\frac{c_tw_{i_t,\tau_t-1}^2 - w_{i_t,\tau_t}^2}{c_tw_{i_t,\tau_t}^2w_{i_t,\tau_t-1}^2}\right)\widehat{\varphi}_t'\widetilde{\beta}_{t-1}\varepsilon_t\right| \\ &\quad + 2\left|\left(\frac{c_tw_{i_t,\tau_t-1}^2 - w_{i_t,\tau_t}^2}{c_tw_{i_t,\tau_t}^2w_{i_t,\tau_t-1}^2}\right)\widehat{\varphi}_t'\widehat{R}_t^{-1}\widehat{\varphi}_t\widehat{\varepsilon}_t\varepsilon_t\right|, \end{aligned} \tag{3.4.30}$$

where according to (3.3.10) the constant c_t is, for convenience, given by

$$c_t = \begin{cases} \dfrac{\tau_t}{\tau_t+1} & \text{if } 0 < \tau_t \leq \delta, \\ \dfrac{\tau_t-\delta}{\tau_t+1-\delta} & \text{if } \quad \tau_t > \delta, \end{cases}$$

so that

$$c_tw_{i_t,\tau_t-1}^2 - w_{i_t,\tau_t}^2 = \begin{cases} -\dfrac{1}{\tau_t+1}\widehat{\varepsilon}_t^2 \\ -\dfrac{1}{\tau_t+1-\delta}\widehat{\varepsilon}_t^2 \end{cases} = -\frac{k_S}{t}\widehat{\varepsilon}_t^2,$$

for some suitable bounded function $k_S = k(S)$ of S, where $\tau_t = \lceil \frac{t-1}{S} \rceil$. Since $w_{i,\tau}^2 = \widehat{\sigma}_i^{2(i,\tau)}$ is a strongly consistent estimator for $\sigma_i^2 > 0$, then $\dfrac{1}{c_t w_{i_t,\tau_t}^2 w_{i_t,\tau_t-1}^2}$ is bounded $a.s.$ Let M be such a bounding constant. Hence, from (3.4.30) we obtain

$$\begin{aligned} T_t \le T_{t-1} - 2a_t b_t + 2c_t^{-1} w_{i_t,\tau_t-1}^{-2} \widehat{\varphi}_t' \widetilde{\beta}_{t-1} \varepsilon_t + 2c_t^{-1} w_{i_t,\tau_t-1}^{-2} \widehat{\varphi}_t' \widehat{R}_t^{-1} \widehat{\varphi}_t \left(\widehat{\varepsilon}_t - \varepsilon_t\right) \varepsilon_t \\ + 2w_{i_t,\tau_t-1}^{-2} \widehat{\varphi}_t' \widehat{R}_t^{-1} \widehat{\varphi}_t \varepsilon_t^2 + \frac{2k_S M}{t} \left| \widehat{\varphi}_t' \widetilde{\beta}_{t-1} \widehat{\varepsilon}_t^2 \varepsilon_t \right| + 2\frac{2k_S M}{t} \left| \widehat{\varphi}_t' \widehat{R}_t^{-1} \widehat{\varphi}_t \widehat{\varepsilon}_t^3 \varepsilon_t \right| \end{aligned} \quad a.s. \qquad (3.4.31a)$$

As the quantities w_{i_t,τ_t-1}, $\widehat{\varphi}_t$, $\widehat{R}_t$ and $\widetilde{\beta}_{t-1}$ are $\mathcal{F}_{t-1}$-measurable and since $E(\varepsilon_t/\mathcal{F}_{t-1}) = 0$ and $E(\varepsilon_t^2/\mathcal{F}_{t-1}) = \sigma_t^2$, taking the conditional expectation of both sides of (3.4.31a) with respect to $\mathcal{F}_{t-1}$ we then get

$$\begin{aligned} E(T_t + 2a_t b_t/\mathcal{F}_{t-1}) \le T_{t-1} + 2c_t^{-1} w_{i_t,\tau_t-1}^{-2} \widehat{\varphi}_t' \widehat{R}_t^{-1} \widehat{\varphi}_t \sigma_t^2 + \frac{2k_S M}{t} \left| \widehat{\varphi}_t' \widetilde{\beta}_{t-1} \right| E\left(\left| \widehat{\varepsilon}_t^2 \varepsilon_t \right| / \mathcal{F}_{t-1} \right) \\ +2\frac{2k_S M}{t} \widehat{\varphi}_t' \widehat{R}_t^{-1} \widehat{\varphi}_t E\left(\left| \widehat{\varepsilon}_t^3 \varepsilon_t \right| / \mathcal{F}_{t-1} \right). \end{aligned}$$

Add $2\sum_{k=1}^{t-1} a_k b_k$ to both sides of this last equation. This gives

$$\begin{aligned} E\left(t\overline{T}_t / \mathcal{F}_{t-1} \right) \le (t-1)\overline{T}_{t-1} + 2c_t^{-1} w_{i_t,\tau_t-1}^{-2} \widehat{\varphi}_t' \widehat{R}_t^{-1} \widehat{\varphi}_t \sigma_t^2 + \frac{2k_S M}{t} \left| \widehat{\varphi}_t' \widetilde{\beta}_{t-1} \right| \times \\ E\left(\left| \widehat{\varepsilon}_t^2 \varepsilon_t \right| / \mathcal{F}_{t-1} \right) + \frac{2k_S M}{t} \widehat{\varphi}_t' \widehat{R}_t^{-1} \widehat{\varphi}_t E\left(\left| \widehat{\varepsilon}_t^3 \varepsilon_t \right| / \mathcal{F}_{t-1} \right). \end{aligned} \qquad (3.4.31b)$$

Hence, after some simple manipulations, with the fact that $\widehat{\varphi}_t' \widetilde{\beta}_{t-1} = y_t - \widehat{\varepsilon}_t$ and $C = 2k_S M$, we finally find (3.4.29), which proves the lemma. ∎

Now, we will apply Robbins and Siegmund's (1971) Martingale result given by Lemma 3.4.1 to (3.4.31a). We have only to show that $\overline{T}_{t-1}$ is positive and that the two last terms in (3.4.31a) are summable, i.e.

$$\frac{1}{t}\left(T_t + 2\sum_{k=1}^{t} a_k b_k \right) \ge 0, \qquad (3.4.32a)$$

$$\sum \frac{2}{t} c_t^{-1} w_{i_t,\tau_t-1}^{-2} \widehat{\varphi}_t' \widehat{R}_t^{-1} \widehat{\varphi}_t \sigma_t^2 < \infty, \qquad (3.4.32b)$$

$$\sum C t^{-2} \left[\left| y_t - \widehat{\varepsilon}_t \right| E\left(\left| \widehat{\varepsilon}_t^2 \varepsilon_t \right| / \mathcal{F}_{t-1} \right) + \widehat{\varphi}_t' \widehat{R}_t^{-1} \widehat{\varphi}_t E\left(\left| \widehat{\varepsilon}_t^3 \varepsilon_t \right| / \mathcal{F}_{t-1} \right) \right] < \infty. \qquad (3.4.32c)$$

The remaining lemmas deal with the proof of the relations (3.4.32).

It should be noted that equation (3.4.32a) holds once the model linking a_k to b_k given below (equation (3.4.34)) is very strictly passive (Hill and Moylan, 1980), i.e. there exist μ, $\eta > 0$ and $K \geq 0$ such that

$$\sum_{t=1}^{T} \left(a_t b_t - \mu a_t^2 - \eta b_t^2\right) + K \geq 0 \quad a.s. \quad \forall T \geq 1. \tag{3.4.33}$$

A frequency domain sufficient condition for (3.4.33) to hold is given in Adams and Goodwin (1995). This condition is intractable in practice and will not be considered here. A more simple and weaker condition is given by (3.4.12) (assumption $B2$). In the following lemma we show that assumption $B2$ assures the positivity of $\sum_{k=1}^{t} a_k b_k$ and hence that of $\overline{T}_t$. Before giving this lemma, we need to quantify the relationship between a_k and b_k which is a key step in the proof of the theorem.

Lemma 3.4.4

With a_t and b_t given in (3.4.26) *we have*

$$\begin{aligned} b_t &= \left(\theta_{i_t}(L)^{-1} - 1/2\right) a_t, \qquad (3.4.34) \\ &= H_{i_t}(L) a_t, \\ &= \left(\sum_{s=1}^{\infty} H_{i_t,s} L^s\right) a_t, \end{aligned}$$

where $\theta_i(L) = 1 - \sum_{j=1}^{p} \theta_{i,j} L^j$ *and* $i_t \in \{1, ..., S\}$ *is such that* $t = i_t + S\tau_t$.

Proof

We have

$$\begin{aligned} \widehat{\epsilon}_t - \varepsilon_t &= y_t - \widehat{\varphi}_t' \widehat{\beta}_t - y_t + \varphi_t' \beta^*, \\ &= \varphi_t' \beta^* - \widehat{\varphi}_t' \widehat{\beta}_t + \widehat{\varphi}_t' \beta^* - \widehat{\varphi}_t' \beta^*, \\ &= -\widehat{\varphi}_t' \widetilde{\beta}_t + (\varphi_t - \widehat{\varphi}_t)' \beta^*, \\ &= -\widehat{\varphi}_t' \widetilde{\beta}_t + \sum_{j=1}^{p} \theta_{i_t,j} (\widehat{\epsilon}_{t-j} - \varepsilon_{t-j}). \qquad (3.4.35) \end{aligned}$$

Thus, from (3.4.35) we obtain

$$\begin{aligned}\widehat{\epsilon}_t - \varepsilon_t &= -\theta_{i_t}(L)^{-1}\widehat{\varphi}_t'\widetilde{\beta}_t, \\ &= w_{i_t,\tau_t}\theta_{i_t}(L)^{-1}a_t, \end{aligned} \tag{3.4.36}$$

provided that $\theta_{i_t}(L)$ is stable. Combining (3.4.26b) and (3.4.36), one get (3.4.34). ■

Lemma 3.4.5

Provided that $B2$ holds, we have

$$\sum_{k=1}^{t} a_k b_k \geq \lambda \sum_{k=1}^{t} a_k^2, \text{ for any } t \geq 1, \tag{3.4.37}$$

where

$$\lambda = \inf_{s\in\{1,..,S\}} \inf_{-\pi\leq\omega\leq\pi} \left[\operatorname{Re} H_s(e^{i\omega})\right].$$

Proof

From the definition of a_k and b_k, we can write

$$\begin{aligned}\sum_{k=1}^{t} a_k b_k &= \sum_{k=1}^{t} a_k \left(\sum_{j=0}^{\infty} H_{k,j}L^j\right) a_k, \\ &= \sum_{k=1}^{t} a_k \sum_{j=0}^{\infty} H_{k,j}a_{k-j}, \\ &= \sum_{j=0}^{\infty}\sum_{k=1}^{t} H_{k,j}a_k a_{k-j}. \end{aligned} \tag{3.4.38}$$

For convenience and without loss of generality, suppose that t is a multiple of S, that is $t = S\tau$. Then we can rewrite (3.4.38)

$$\sum_{k=1}^{S\tau} a_k b_k = \sum_{j=0}^{\infty}\sum_{s=1}^{S} H_{s,j} \sum_{m=0}^{\tau-1} a_{s+Sm}a_{s+Sm-j}. \tag{3.4.39}$$

Using the facts that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{ik\omega}d\omega = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\sum_{m=0}^{\tau-1} a_{s+Sm}a_{s+Sm-j} = \sum\sum_{\{n,l\in\{0,\ldots\tau-1\}/Sl-Sn+j=0\}} a_{s+Sl}a_{s+Sn},$$

we have

$$\begin{aligned}
\sum_{k=1}^{S\tau} a_k b_k &= \sum_{j=0}^{\infty}\sum_{s=1}^{S} H_{s,j}\sum_{l=0}^{\tau-1}\sum_{n=0}^{\tau-1}\frac{1}{2\pi}\int_{-\pi}^{\pi} a_{s+Sl}a_{s+Sn}e^{i(Sl-Sn+j)\omega}d\omega, \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{j=0}^{\infty}\sum_{s=1}^{S} H_{s,j}\left(\sum_{l=0}^{\tau-1} a_{s+Sl}e^{iSl\omega}\right)\left(\sum_{n=0}^{\tau-1} a_{s+Sn}e^{-iSn\omega}\right)e^{ij\omega}d\omega, \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{j=0}^{\infty}\sum_{s=1}^{S} H_{s,j}\left|\sum_{l=0}^{\tau-1} a_{s+Sl}e^{iSl\omega}\right|^2 e^{ij\omega}d\omega, \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{s=1}^{S}\left|\sum_{l=0}^{\tau-1} a_{s+Sl}e^{iSl\omega}\right|^2\left(\sum_{j=0}^{\infty} H_{s,j}e^{ij\omega}\right)d\omega, \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{s=1}^{S}\left|\sum_{l=0}^{\tau-1} a_{s+Sl}e^{iSl\omega}\right|^2 H_s(e^{i\omega})d\omega, \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{s=1}^{S}\left|\sum_{l=0}^{\tau-1} a_{s+Sl}e^{iSl\omega}\right|^2 \operatorname{Re} H_s(e^{i\omega})d\omega \\
&\geq \lambda\sum_{s=1}^{S}\frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\sum_{l=0}^{\tau-1} a_{s+Sl}e^{iSl\omega}\right|^2 d\omega \\
&= \lambda\sum_{s=1}^{S}\sum_{l=0}^{\tau-1} a_{s+Sl}^2, \\
&= \lambda\sum_{k=1}^{S\tau} a_k^2,
\end{aligned}$$

where $\lambda = \inf_{s\in\{1,..,S\}}\inf_{-\pi\leq\omega\leq\pi}\left[\operatorname{Re} H_s(e^{i\omega})\right]$. Using the real positiveness assumption $B2$, we get from this last inequality

$$\begin{aligned}\sum_{k=1}^{S\tau} a_k b_k &\geq \lambda \sum_{k=1}^{S\tau} a_k^2, \\ &\geq 0.\end{aligned}$$

This completes the proof. ■

To establish the proof of the theorem, we need to prove (3.4.32b). The following lemma, proves an equivalent relation to (3.4.32b) under a certain assumption.

Lemma 3.4.6

Provided that the following condition

$$\lim_{t \longrightarrow \infty} \sup t^{-1} \sum_{k=1}^{t} w_{i_t,\tau_t}^{-2} \left\| \widehat{\varphi}_k \right\|^2 < \infty, \tag{3.4.40}$$

holds, we have

$$\sum_{t \geq 1} \frac{1}{t} w_{i_t,\tau}^{-2} \widehat{\varphi}_t' \widehat{R}_t^{-1} \widehat{\varphi}_t < \infty \text{ , } a.s. \tag{3.4.41}$$

Proof

First, we have

$$\begin{aligned}\widehat{\varphi}_t' \widehat{R}_t^{-1} \widehat{\varphi}_t &= tr\left(\widehat{\varphi}_t' \widehat{R}_t^{-1} \widehat{\varphi}_t\right), \\ &= tr\left(\widehat{\varphi}_t' \left(\widehat{R}_t^{-1}\right)^2 \widehat{R}_t \widehat{\varphi}_t\right), \\ &= tr\left(\widehat{\varphi}_t \widehat{\varphi}_t' \left(\widehat{R}_t^{-1}\right)^2 \widehat{R}_t\right), \\ &\leq tr\left(\widehat{\varphi}_t \widehat{\varphi}_t' \left(\widehat{R}_t^{-1}\right)^2\right) tr\left(\widehat{R}_t\right), \\ &= \widehat{\varphi}_t' \left(\widehat{R}_t^{-1}\right)^2 \widehat{\varphi}_t tr\left(\widehat{R}_t\right), \\ &= \widehat{\varphi}_t' \left(\widehat{R}_t^{-1}\right)^2 \widehat{\varphi}_t \sum_{k=1}^{t} w_{i_t,\tau}^{-2} \left\| \widehat{\varphi}_k \right\|^2.\end{aligned}$$

In the latter inequality we have used the well-known trace inequality for nonnegative matrices (e.g. Yang and Feng, 2002, Lemma 2.3)

$$tr\,(AB) \leq tr\,(A)\,tr\,(B)\,, \text{ for any nonnegative matrices } A \text{ and } B$$

By using the properties of the matrix inversion lemma given in Appendix A, we can write

$$\begin{aligned}
\sum_{t=1}^{\infty} t^{-1} w_{i_t,\tau}^{-2} \widehat{\varphi}_t' \widehat{R}_t^{-1} \widehat{\varphi}_t &\leq \sum_{t=1}^{\infty} w_{i_t,\tau}^{-2} \widehat{\varphi}_t' \left(\widehat{R}_t^{-1}\right)^2 \widehat{\varphi}_t \left(t^{-1} \sum_{k=1}^{t} w_{i_t,\tau}^{-2} \left\|\widehat{\varphi}_k\right\|^2\right), \\
&= \sum_{t=1}^{\infty} \widehat{\varphi}_t' \widehat{R}_t^{-1} \frac{\widehat{R}_{t-1}^{-1} \widehat{\varphi}_t}{w_{i_t,\tau}^2 + \widehat{\varphi}_t' \widehat{R}_{t-1}^{-1} \widehat{\varphi}_t} \left(t^{-1} \sum_{k=1}^{t} w_{i_t,\tau}^{-2} \left\|\widehat{\varphi}_k\right\|^2\right), \\
&\leq \sum_{t=1}^{\infty} w_{i_t,\tau}^{-2} \widehat{\varphi}_t' \widehat{R}_t^{-1} \widehat{R}_{t-1}^{-1} \widehat{\varphi}_t \left(t^{-1} \sum_{k=1}^{t} w_{i_t,\tau}^{-2} \left\|\widehat{\varphi}_k\right\|^2\right), \\
&= \sum_{t=1}^{\infty} tr\left(w_{i_t,\tau}^{-2} \widehat{\varphi}_t \widehat{\varphi}_t' \widehat{R}_t^{-1} \widehat{R}_{t-1}^{-1}\right) \left(t^{-1} \sum_{k=1}^{t} w_{i_t,\tau}^{-2} \left\|\widehat{\varphi}_k\right\|^2\right), \\
&= \sum_{t=1}^{\infty} tr\left(\left(\widehat{R}_t - \widehat{R}_{t-1}\right) \widehat{R}_t^{-1} \widehat{R}_{t-1}^{-1}\right) \left(t^{-1} \sum_{k=1}^{t} w_{i_t,\tau}^{-2} \left\|\widehat{\varphi}_k\right\|^2\right), \\
&= \sum_{t=1}^{\infty} tr\left(\widehat{R}_{t-1}^{-1} - \widehat{R}_t^{-1}\right) \left(t^{-1} \sum_{k=1}^{t} w_{i_t,\tau}^{-2} \left\|\widehat{\varphi}_k\right\|^2\right),
\end{aligned}$$

since

$$\sum_{t=1}^{\infty} tr\left(\widehat{R}_{t-1}^{-1} - \widehat{R}_t^{-1}\right) = tr(\widehat{R}_0^{-1}) - tr(\widehat{R}_\infty^{-1}) < \infty,$$

by the assumption (3.4.40) we can conclude that (3.4.41) is true, which completes the proof. ■

Remark 3.4.6

A strong counterpart of Theorem 1.6.2 (equation (1.6.10)), that is

$$w_{i,\tau}^2 = \widehat{\sigma}_i^{2(i,\tau)} \overset{a.s.}{\rightarrow} \sigma_i^2 > 0 \text{ as } \tau \to \infty,$$

can be easily proved. Therefore $w_{i,\tau}^{-2}$ and $w_{i,\tau-1}^{-2}$ are bounded $a.s.$ so that (3.4.32b) and (3.4.41) may be equivalent.

It is worth noting that the assumption (3.4.40) is intractable as it stands. The following lemma shows that assumption $B1a$ implies (3.4.40).

Lemma 3.4.7

If the assumption $B1a$ holds, which implies

$$\lim_{t\to\infty} \sup \frac{1}{t} \sum_{k=1}^{t} \frac{1}{\sigma_k^2} y_k^2 < \infty, \tag{3.4.42a}$$

then (3.4.40) *is satisfied.*

Proof

By similar arguments as for the proof of relation (3.3.15), one can easily show under $B1$ that

$$\frac{1}{t} \sum_{k=1}^{t} \left(w_{i_k,\tau_k}^{-2} - \sigma_k^{-2} \right) y_k^2 = o_p(1).$$

This immediately implies that

$$\lim_{t\to\infty} \sup \frac{1}{t} \sum_{k=1}^{t} w_{i_k,\tau_k}^{-2} y_k^2 < \infty. \tag{3.4.42b}$$

Hence to establish the lemma, it remains to show that (3.4.42a) implies

$$\lim_{t\longrightarrow\infty} \sup \frac{1}{t} \sum_{k=1}^{t} w_{i_k,\tau_k}^{-2} \widehat{\epsilon}_k^2 < \infty. \tag{3.4.43}$$

To do this, premultiplying (3.4.18b) by $\widehat{R}_t$ and the resulting equation by $\widehat{\beta}_t'$, after some simple manipulations one gets

$$\begin{aligned}
\widehat{\beta}_t' \widehat{R}_t \widehat{\beta}_t &= \widehat{\beta}_t' \widehat{R}_t \widehat{\beta}_{t-1} + w_{i_t,\tau_t}^{-2} \widehat{\beta}_t' \widehat{\varphi}_t \widehat{\varepsilon}_t, \\
&= \left(\widehat{\beta}_{t-1} + w_{i_t,\tau_t}^{-2} \widehat{R}_t^{-1} \widehat{\varphi}_t \widehat{\varepsilon}_t \right)' \widehat{R}_t \widehat{\beta}_{t-1} + w_{i_t,\tau_t}^{-2} \widehat{\beta}_t' \widehat{\varphi}_t \widehat{\varepsilon}_t, \\
&= \widehat{\beta}_{t-1}' \widehat{R}_t \widehat{\beta}_{t-1} + w_{i_t,\tau_t}^{-2} \widehat{\varphi}_t' \widehat{\beta}_{t-1} \widehat{\varepsilon}_t + w_{i_t,\tau_t}^{-2} \widehat{\beta}_t' \widehat{\varphi}_t \widehat{\varepsilon}_t, \\
&= \begin{array}{r} \widehat{\beta}_{t-1}' \left(\widehat{R}_{t-1} + w_{i_t,\tau_t}^{-2} \widehat{\varphi}_t \widehat{\varphi}_t' \right) \widehat{\beta}_{t-1} + w_{i_t,\tau_t}^{-2} \widehat{\varphi}_t' \widehat{\beta}_{t-1} \widehat{\varepsilon}_t + \\ w_{i_t,\tau_t}^{-2} \left(\widehat{\beta}_{t-1} + w_{i_t,\tau_t}^{-2} \widehat{R}_t^{-1} \widehat{\varphi}_t \widehat{\varepsilon}_t \right)' \widehat{\varphi}_t \widehat{\varepsilon}_t, \end{array} \\
&= \begin{array}{r} \widehat{\beta}_{t-1}' \widehat{R}_{t-1} \widehat{\beta}_{t-1} + w_{i_t,\tau_t}^{-2} \left(\widehat{\varphi}_t' \widehat{\beta}_{t-1} \right)^2 + 2 w_{i_t,\tau_t}^{-2} \widehat{\varphi}_t' \widehat{\beta}_{t-1} \widehat{\varepsilon}_t + \\ w_{i_t,\tau_t}^{-4} \widehat{\varphi}_t' \widehat{R}_t^{-1} \widehat{\varphi}_t \widehat{\varepsilon}_t^2. \end{array}
\end{aligned} \tag{3.4.44}$$

Knowing that $\widehat{\varphi}_t'\widehat{\beta}_{t-1} = y_t - \widehat{\varepsilon}_t$, equation (3.4.44) can be written

$$\begin{aligned}
\widehat{\beta}_t'\widehat{R}_t\widehat{\beta}_t &= \widehat{\beta}_{t-1}'\widehat{R}_{t-1}\widehat{\beta}_{t-1} + w_{i_t,\tau_t}^{-2}y_t^2 + w_{i_t,\tau_t}^{-2}\widehat{\varepsilon}_t^2 - 2w_{i_t,\tau_t}^{-2}y_t\widehat{\varepsilon}_t + 2w_{i_t,\tau_t}^{-2}\widehat{\varphi}_t'\widehat{\beta}_{t-1}\widehat{\varepsilon}_t \\
&\qquad + w_{i_t,\tau_t}^{-4}\widehat{\varphi}_t'\widehat{R}_t^{-1}\widehat{\varphi}_t\widehat{\varepsilon}_t^2, \\
&= \widehat{\beta}_{t-1}'\widehat{R}_{t-1}\widehat{\beta}_{t-1} + w_{i_t,\tau_t}^{-2}y_t^2 + w_{i_t,\tau_t}^{-2}\widehat{\varepsilon}_t^2 - 2w_{i_t,\tau_t}^{-2}\left(\widehat{\varphi}_t'\widehat{\beta}_{t-1} + \widehat{\varepsilon}_t\right)\widehat{\varepsilon}_t \\
&\qquad + 2w_{i_t,\tau_t}^{-2}\widehat{\varphi}_t'\widehat{\beta}_{t-1}\widehat{\varepsilon}_t + w_{i_t,\tau_t}^{-4}\widehat{\varphi}_t'\widehat{R}_t^{-1}\widehat{\varphi}_t\widehat{\varepsilon}_t^2, \\
&= \widehat{\beta}_{t-1}'\widehat{R}_{t-1}\widehat{\beta}_{t-1} + w_{i_t,\tau_t}^{-2}y_t^2 + w_{i_t,\tau_t}^{-2}\widehat{\varepsilon}_t^2 - w_{i_t,\tau_t}^{-2}\widehat{\varepsilon}_t^2\left(1 - w_{i_t,\tau_t}^{-2}\widehat{\varphi}_t'\widehat{R}_t^{-1}\widehat{\varphi}_t\right). \qquad (3.4.45)
\end{aligned}$$

Summing the two sides of (3.4.45) for $k = 1, ..., t$, and using equation (3.4.20) we obtain

$$\begin{aligned}
\widehat{\beta}_t'\widehat{R}_t\widehat{\beta}_t &= \widehat{\beta}_0'\widehat{R}_0\widehat{\beta}_0 + \sum_{k=1}^{t} w_{i_k,\tau_k}^{-2}y_k^2 - \sum_{k=1}^{t} w_{i_k,\tau_k}^{-2}\widehat{\varepsilon}_k^2\left(1 - w_{i_k,\tau_k}^{-2}\widehat{\varphi}_k'\widehat{R}_k^{-1}\widehat{\varphi}_k\right). \\
&= \widehat{\beta}_0'\widehat{R}_0\widehat{\beta}_0 + \sum_{k=1}^{t} w_{i_k,\tau_k}^{-2}y_k^2 + \sum_{k=1}^{t}\frac{w_{i_k,\tau_k}^{-2}\widehat{\epsilon}_k^2}{1 - w_{i_k,\tau_k}^{-2}\widehat{\varphi}_k'\widehat{R}_k^{-1}\widehat{\varphi}_k}.
\end{aligned}$$

Hence

$$\frac{1}{t}\sum_{k=1}^{t}\frac{w_{i_k,\tau_k}^{-2}\widehat{\epsilon}_k^2}{1 - w_{i_k,\tau_k}^{-2}\widehat{\varphi}_k'\widehat{R}_k^{-1}\widehat{\varphi}_k} = \frac{1}{t}\sum_{k=1}^{t} w_{i_k,\tau_k}^{-2}y_k^2 + \frac{1}{t}\widehat{\beta}_0'\widehat{R}_0\widehat{\beta}_0 - \frac{1}{t}\widehat{\beta}_t'\widehat{R}_t\widehat{\beta}_t. \qquad (3.4.46)$$

Thus, from (3.4.46) it is clear that (3.4.42b) implies

$$\lim_{t\longrightarrow\infty}\sup\frac{1}{t}\sum_{k=1}^{t}\frac{w_{i_k,\tau_k}^{-2}\widehat{\epsilon}_k^2}{1 - w_{i_k,\tau_k}^{-2}\widehat{\varphi}_k'\widehat{R}_k^{-1}\widehat{\varphi}_k} < \infty,$$

and as $w_{i_k,\tau_k}^{-2}\widehat{\varphi}_k'\widehat{R}_k^{-1}\widehat{\varphi}_k < 1$ (see Appendix A equation ($A4$)) we then have

$$\frac{w_{i_k,\tau_k}^{-2}\widehat{\epsilon}_k^2}{1 - w_{i_k,\tau_k}^{-2}\widehat{\varphi}_k'\widehat{R}_k^{-1}\widehat{\varphi}_k} \geq w_{i_k,\tau_k}^{-2}\widehat{\epsilon}_k^2,$$

which implies (3.4.43) and hence (3.4.40). This completes the proof. ■

Lemma 3.4.8

Under assumptions $B1$ and $B2$ we have

$$\sum_{t\geq 1} t^{-2}\widehat{\varphi}_t'\widehat{R}_t^{-1}\widehat{\varphi}_t E\left(\left|\widehat{\varepsilon}_t^3\varepsilon_t\right|/\mathcal{F}_{t-1}\right) < \infty \ a.s. \qquad (3.4.47a)$$

$$\sum_{t\geq 1} t^{-2}\left|y_t - \widehat{\varepsilon}_t\right| E\left(\left|\widehat{\varepsilon}_t^2\varepsilon_t\right|/\mathcal{F}_{t-1}\right) < \infty \ a.s. \qquad (3.4.47b)$$

Proof

From Lemma 3.4.6 and the boundedness of $\widehat{\sigma}_k^{2(i_k,\tau_k)}$ (see Remark 3.4.6), we have

$$\sum_{t\geq 1} t^{-1}\widehat{\varphi}_t'\widehat{R}_t^{-1}\widehat{\varphi}_t < \infty, \quad a.s.$$

Thus to establish (3.4.47a), it suffices to show that $t^{-1}E\left(\left|\widehat{\varepsilon}_t^3\varepsilon_t\right|/\mathcal{F}_{t-1}\right)$ is bounded $a.s.$ Indeed, by the Markov inequality and assumptions $B2b$ and $B4$ (see also Section 3.5 for more details) we have

$$\begin{aligned} P\left(t^{-1}E\left(\left|\widehat{\varepsilon}_t^3\varepsilon_t\right|/\mathcal{F}_{t-1}\right) \geq \eta\right) &\leq \eta^{-1}t^{-1}E\left[E\left(\left|\widehat{\varepsilon}_t^3\varepsilon_t\right|/\mathcal{F}_{t-1}\right)\right], \\ &= \eta^{-1}t^{-1}E\left(\left|\widehat{\varepsilon}_t^3\varepsilon_t\right|\right), \\ &\leq Ct^{-1}E\left(\left|\varepsilon_t^4\right|\right), \\ &\overset{t\to\infty}{\to} 0, \end{aligned}$$

for any $\eta > 0$, where C is an appropriate positive constant. Thus, (3.4.47a) is established. Returning now to (3.4.47b). Using again the Markov inequality and assumptions $B2$ we have

$$\begin{aligned} P\left(t^{-\delta}\left|y_t - \widehat{\varepsilon}_t\right| E\left(\left|\widehat{\varepsilon}_t^2\varepsilon_t\right|/\mathcal{F}_{t-1}\right) \geq \eta\right) &= P\left(t^{-\delta}E\left(\left|\left(y_t - \widehat{\varepsilon}_t\right)\widehat{\varepsilon}_t^2\varepsilon_t\right|/\mathcal{F}_{t-1}\right) \geq \eta\right), \\ &\leq \eta^{-1}t^{-\delta}E\left(\left(y_t - \widehat{\varepsilon}_t\right)\widehat{\varepsilon}_t^2\varepsilon_t\right), \\ &\overset{t\to\infty}{\to} 0, \end{aligned}$$

provided that the model linking y_t to ε_t is causal. Hence, $t^{-\delta}\left|y_t - \widehat{\varepsilon}_t\right| E\left(\left|\widehat{\varepsilon}_t^2\varepsilon_t\right|/\mathcal{F}_{t-1}\right)$ is bounded $a.s.$ and then the sum

$$\sum_{t\geq 1} t^{-2+\delta}\left[t^{-\delta}\left|y_t - \widehat{\varepsilon}_t\right| E\left(\left|\widehat{\varepsilon}_t^2\varepsilon_t\right|/\mathcal{F}_{t-1}\right)\right] = \sum_{t\geq 1} t^{-2}\left|y_t - \widehat{\varepsilon}_t\right| E\left(\left|\widehat{\varepsilon}_t^2\varepsilon_t\right|/\mathcal{F}_{t-1}\right),$$

is finite $a.s.$, which achieves the proof. ■

Proof of Theorem 3.4.1

i) We shall apply the Robbins-Siegmund Martingale lemma (Lemma 3.4.1) to the function

$$\overline{T}_t = \frac{1}{t}\left(T_t + 2\sum_{k=1}^{t} a_k b_k\right),$$

with L_t, β_t and α_t in Lemma 3.4.1 corresponding to

$$\begin{aligned} L_t &= \overline{T}_t, \\ \beta_t &= t^{-1}\overline{T}_{t-1}, \\ \alpha_t &= Ct^{-2}\left[|y_t - \widehat{\varepsilon}_t|\, E\left(\left|\widehat{\varepsilon}_t^2\varepsilon_t\right|/\mathcal{F}_{t-1}\right) + \widehat{\varphi}_t'\widehat{R}_t^{-1}\widehat{\varphi}_t E\left(\left|\widehat{\varepsilon}_t^3\varepsilon_t\right|/\mathcal{F}_{t-1}\right)\right]. \end{aligned}$$

By Lemma 3.4.5, $t^{-1}\overline{T}_{t-1}$ is positive and by Lemma 3.4.6 and Lemma 3.4.8, α_t is absolutely summable. Therefore, the conclusions of Lemma 3.4.1 are true, i.e. there exists a nonnegative finite random variable $\overline{T}$ such that

$$\overline{T}_t \overset{t\to\infty}{\to} \overline{T} \quad a.s.,$$

and

$$\sum_{t=1}^{\infty} t^{-1}\overline{T}_{t-1} < \infty \quad a.s. \tag{3.4.48}$$

The latter conclusion (3.4.48) implies necessarily that

$$\overline{T} = 0,\ a.s.$$

Since $\overline{T}_t$ is the sum of two positive terms we conclude that

$$t^{-1}\left(\widehat{\beta}_t - \beta^*\right)'\widehat{R}_t\left(\widehat{\beta}_t - \beta^*\right) \overset{t\to\infty}{\to} 0,\ a.s.,$$

$$t^{-1}\sum_{k=1}^{t} a_k b_k \overset{t\to\infty}{\to} 0,\ a.s. \tag{3.4.49}$$

Thus the conclusion (3.4.13a) of Theorem 3.4.1 has been proved. To prove (3.4.13b) we note that Lemma 3.4.5 implies that

$$\sum_{k=1}^{t} a_k^2 \overset{t\to\infty}{\to} 0,\ a.s., \tag{3.4.50a}$$

and hence

$$\sum_{k=1}^{t} b_k^2 \overset{t\to\infty}{\to} 0,\ a.s., \tag{3.4.50b}$$

since b_t is obtained by exponentially stable filtering of a_t. Moreover, as we have from (3.4.26)

$$\widehat{\epsilon}_t - \varepsilon_t = w_{i_t,\tau_t} b_t - \frac{1}{2} w_{i_t,\tau_t} a_t,$$

the conclusion (3.4.13b) of Theorem 3.4.1 follows from (3.4.49) and (3.4.50). The first part of Theorem 3.4.1 is now proved.

ii) To show the second part of the theorem, note that under $B3$ we can write

$$\begin{aligned}\left(\frac{1}{t}\widehat{R}_t - R\right) &= \frac{1}{t}\sum_{k=1}^{t}\left(\frac{1}{\widehat{\sigma}_k^{2(i_k,\tau_k)}}\widehat{\varphi}_k\widehat{\varphi}_k' - \frac{1}{\sigma_k^2}\varphi_k\varphi_k'\right), \\ &= \frac{1}{t}\sum_{k=1}^{t}\left(\frac{1}{\widehat{\sigma}_k^{2(i_k,\tau_k)}} - \frac{1}{\sigma_k^2}\right)\widehat{\varphi}_k\widehat{\varphi}_k' + \frac{1}{t}\sum_{k=1}^{t}\frac{1}{\sigma_k^2}\left(\varphi_k\varphi_k' - \widehat{\varphi}_k\widehat{\varphi}_k'\right). \qquad (3.4.51)\end{aligned}$$

Using similar techniques for proving (3.3.15) (cf, Section 3.3), one can easily show that, under the existence of fourth moment ($B2b$), the first term in the right member of (3.4.51) can be neglected, i.e.

$$\frac{1}{t}\sum_{k=1}^{t}\left(\frac{1}{\widehat{\sigma}_k^{2(i_k,\tau_k)}} - \frac{1}{\sigma_k^2}\right)\widehat{\varphi}_k\widehat{\varphi}_k' = o_p(1).$$

Moreover, by the conclusion (3.4.13b) of the first part of Theorem 3.4.1, it also follows that

$$\frac{1}{t}\sum_{k=1}^{t}\frac{1}{\sigma_k^2}\left(\varphi_k\varphi_k' - \widehat{\varphi}_k\widehat{\varphi}_k'\right) = o_p(1),$$

and then

$$\left(\frac{1}{t}\widehat{R}_t - R\right) = o_p(1). \qquad (3.4.52)$$

Therefore, by $B3$ and (3.4.52) we conclude that $\frac{1}{t}\widehat{R}_t$ is bounded by a positive definite matrix and hence, from the conclusion (3.4.13a) of Theorem 3.4.1, we finally obtain

$$\widehat{\beta}_t \overset{t\to\infty}{\to} \beta^*,$$

which completes the Proof of part ii).

iii) When we use in algorithm (3.4.10) the weights $\widehat{\sigma}_{i_t}^{-2(i_t,\tau_t-1)}$ in lieu of $\widehat{\sigma}_{i_t}^{2(i_t,\tau_t)}$ the analysis will be much simpler because of the $\mathcal{F}_{t-1}$-measurability of $\widehat{\sigma}_{i_t}^{-2(i_t,\tau_t-1)}$. Indeed rewriting (3.4.17) as follows

$$T_t \leq T_{t-1} - 2a_tb_t + 2w_{i_t,\tau_t-1}^{-2}\widehat{\varphi}_t'\widetilde{\beta}_{t-1}\varepsilon_t + 2w_{i_t,\tau_t-1}^{-2}\widehat{\varphi}_t'\widehat{R}_t^{-1}\widehat{\varphi}_t\left(\widehat{\varepsilon}_t - \varepsilon_t\right)\varepsilon_t + 2w_{i_t,\tau_t-1}^{-2}\widehat{\varphi}_t'\widehat{R}_t^{-1}\widehat{\varphi}_t\varepsilon_t^2.$$

Now because of the $\mathcal{F}_{t-1}$-measurability of $w^{-2}_{i_t,\tau_t-1}$ Lemma 3.4.3 will be simplified. With $\overline{T}_t = \frac{1}{t}\left(T_t + 2\sum_{k=1}^{t} a_k b_k\right)$, it is easy to show, along the lines of Lemma 3.4.3, that from the $\mathcal{F}_{t-1}$-measurability of $w^{-2}_{i_t,\tau_t-1}$ and $\widehat{\varepsilon}_t - \varepsilon_t$ and the fact that $E(\varepsilon_t^2/\mathcal{F}_{t-1}) = \sigma_t^2$ we have

$$E(\overline{T}_t/\mathcal{F}_{t-1}) \leq \overline{T}_{t-1} - t^{-1}\overline{T}_{t-1} + 2t^{-1}\sigma_t^2 w^{-2}_{i_t,\tau_t-1}\widehat{\varphi}_t'\widehat{R}_t^{-1}\widehat{\varphi}_t \quad a.s.$$

Applying Robbins and Siegmund's (1971) Martingale result given by Lemma 3.4.1 to the latter inequality proves part iii) of the Theorem. ■

Remark 3.4.7

We note that condition (3.4.11c) is used only to do some of the algebraic manipulations. It holds once (3.4.11b) is verified. Otherwise, it is worth mentioning that if we replace the weights w_{i_t,τ_t} by w_{i_t,τ_t-1} in the $PELS$ algorithm (3.4.10), the convergence analysis can be made more easily, since the weights w_{i_t,τ_t-1} will be $\mathcal{F}_{t-1}$-measurable and conditions (3.4.11b) and (3.4.11c) can be omitted. In Chapter 5, we shall return to this important point.

Having shown under $B1 - B4$ the consistency of the $PELS$ estimator, we can in fact also give the rate of convergence. The following Theorem is devoted to this point.

Theorem 3.4.2

i) Under the conditions $B1 - B4$ *of Theorem* 3.4.1*, we have*

$$t^{1/2-\delta}\left(\widehat{\beta}_t - \beta^*\right) \overset{t\to\infty}{\to} 0, \; a.s., \tag{3.4.53}$$

for any $\delta > 0$.

ii) For the $PELS$ *algorithm* (3.4.10) *with the weights* $\widehat{\sigma}_{i_t}^{-2(i_t,\tau_t-1)}$ *instead of* $\widehat{\sigma}_{i_t}^{2(i_t,\tau_t)}$*, the obtained estimator satisfies the conclusion i) without the assumptions* $B1b$ *and* $B4$.

Proof

We only prove part i) of the theorem as the second one is obvious.

By the same approach to prove the latter theorem, we can show (3.4.53), once we replace in

(3.4.31b) the divisor n by n^{δ} for any $\delta > 0$. Thus it suffices to establish the following relations

$$\sum_{t\geq 1} t^{-\delta} w_{i_t,\tau_t-1}^{-2} \widehat{\varphi}_t' \widehat{R}_t^{-1} \widehat{\varphi}_t < \infty, \ a.s., \tag{3.4.54a}$$

$$\sum_{t\geq 1} Ct^{-1-\delta} \left[|y_t - \widehat{\varepsilon}_t| \, E\left(\left| \widehat{\varepsilon}_t^2 \varepsilon_t \right| / \mathcal{F}_{t-1} \right) + \widehat{\varphi}_t' \widehat{R}_t^{-1} \widehat{\varphi}_t E\left(\left| \widehat{\varepsilon}_t^3 \varepsilon_t \right| / \mathcal{F}_{t-1} \right) \right] < \infty \ a.s. \tag{3.4.54b}$$

Relation (3.4.54b) can be easily proved by similar arguments as for (3.4.47) (cf, Lemma 3.4.8). Hence we only prove (3.4.54a) along the lines of Solo (1979). Note that if $\delta \geq 1$, then (3.4.54a) follows immediately from (3.4.47). Next, only the case when $\delta < 1$ will be of interest. Before doing this we need to expose the following lemma due to Solo (1979).

Lemma 3.4.9 (Solo, 1979)

Let (v_t) be a sequence of nondecreasing positive numbers such that

$$\sum_{t=t_0}^{\infty} (v_t - v_{t-1})\, \sigma\left(\widehat{R}_t\right) < \infty,$$

where $\sigma(A)$ denotes the smallest eigenvalue of the matrix A and t_0 is such that $\widehat{R}_t^{-1}$ exists for $t \geq t_0$. Then

$$\sum_{t=t_0}^{\infty} w_{i_t,\tau_t}^{-2} v_{t-1} \widehat{\varphi}_t' \left(\widehat{R}_t^{-1}\right)^2 \widehat{\varphi}_t < \infty.$$

Proof

We refer to Solo (1979, p.962) for the proof. ■

Since $tr\left(\widehat{R}_t\right) = \sum_{k=1}^{t} w_{i_k,\tau_k}^{-2} tr\left(\widehat{\varphi}_k \widehat{\varphi}_k'\right) = \sum_{k=1}^{t} w_{i_k,\tau_k}^{-2} \left\|\widehat{\varphi}_k\right\|^2$ the smallest eigenvalue of

$\widehat{R}_t = \sum_{k=1}^{t} w_{i_k,\tau_k}^{-2} \widehat{\varphi}_k \widehat{\varphi}_k'$ is of order t, that is

$$\frac{\sigma\left(\widehat{R}_t\right)}{t} \sim 1,$$

Thus because of the fact that

$$\sum_{t=1}^{\infty} \frac{(t+1)^{1-\delta} - t^{1-\delta}}{t} < \infty, \text{ for } 0 < \delta < 1,$$

it follows from Lemma 3.4.9 that

$$\sum_{t\geq 1} t^{1-\delta} w_{i_t,\tau_t-1}^{-2} \widehat{\varphi}_t' \widehat{R}_t^{-1} \widehat{\varphi}_t < \infty, \ a.s., \tag{3.4.55}$$

where the nondecreasing sequence (v_t) corresponds to $\left(t^{1-\delta}\right)$. Now since

$$\begin{aligned} \frac{1}{t}\widehat{\varphi}_t' \widehat{R}_t^{-1} \widehat{\varphi}_t &= \frac{1}{t} tr\left(\widehat{\varphi}_t \widehat{\varphi}_t' \left(\widehat{R}_t^{-1}\right)^2 \widehat{R}_t\right), \\ &\leq \frac{1}{t} tr\left(\widehat{\varphi}_t \widehat{\varphi}_t' \left(\widehat{R}_t^{-1}\right)^2\right) tr\left(\widehat{R}_t\right), \\ &= tr\left(\frac{1}{t}\widehat{R}_t\right) \widehat{\varphi}_t' \left(\widehat{R}_t^{-1}\right)^2 \widehat{\varphi}_t, \end{aligned}$$

and by assumption $B3$

$$\lim_{t\to\infty} tr\left(\frac{1}{t}\widehat{R}_t\right) < \infty,$$

and the $a.s.$ boundedness of $\left(w_{i_t,\tau_t}^{-2}\right)$ we conclude that (3.4.54a), which will ensure (3.4.53) and

$$t^{-\delta} \sum_{k=1}^{t} \left(\widehat{\epsilon}_k - \varepsilon_k\right)^2 \xrightarrow{a.s.} 0 \ \text{ as } \ t \to \infty. \tag{3.4.56}$$

■

3.4.3 The $PELS$ algorithm with a prefilter

In the above, we have seen that the $PELS$ algorithm provides consistent estimates provided that the positive realness condition on the transfer functions

$$H_s\left(e^{i\omega}\right) = \left(\frac{1}{\theta_s(e^{i\omega})} - \frac{1}{2}\right), \ s = 1, ..., S,$$

holds. This condition is too restrictive, especially for higher order $PARMA$ models because the class of filters which allow the condition to be satisfied is small. As we have seen in Lemma 3.4.4 and Lemma 3.4.5, the positive real condition results from the particular form of $\widehat{\varphi}_t$ and that of a_t and b_t given by (3.4.26). If we want to relax assumption $B2$, we must replace the pseudo-regressor vector $\widehat{\varphi}_t$ in (3.4.7) by a certain prefiltered vector $\widetilde{\psi}_t$ such that

$$\mu_t(L)\widetilde{\psi}_t = \widehat{\varphi}_t, \tag{3.4.57}$$

where $\mu_t(L) = 1 - \sum_{j=1}^{r} \mu_{tj} L^j$ is a suitable periodic stable filter, i.e. μ_{tj} are periodic in time such that $(\mu_t(L))^{-1}$ exist for any t. The choice of $\mu_t(L)$ is made so that the condition $B2$ will be replaced by a weaker condition as we can see below. Thus we obtain the following modified $PELS$ recursion.

Algorithm 3.4.2 (A modified $PELS$ algorithm with prefiltering)

$$\begin{aligned}
\widehat{\varepsilon}_{i+S\tau} &= y_{i+S\tau} - \widehat{\beta}'_{i+S\tau-1}\widehat{\varphi}_{i+S\tau}, \\
\widehat{\sigma}_i^{2(i,\tau)} &= \begin{cases} \dfrac{\tau}{\tau+1}\,\widehat{\sigma}_i^{2(i,\tau-1)} + \dfrac{1}{\tau+1}\widehat{\varepsilon}^2_{i+S\tau}, \ 0 < \tau \leq \delta, \\ \dfrac{\tau-\delta}{\tau+1-\delta}\,\widehat{\sigma}_i^{2(i,\tau-1)} + \dfrac{1}{\tau+1-\delta}\widehat{\varepsilon}^2_{i+S\tau}, \ \tau > \delta, \end{cases} \\
\widehat{\beta}_{i+S\tau} &= \widehat{\beta}_{i+S\tau-1} + \frac{\widehat{P}_{i+S\tau}\widetilde{\psi}_{i+S\tau}\widehat{\varepsilon}_{i+S\tau}}{\widehat{\sigma}_i^{2(i,\tau)} + \widetilde{\psi}'_{i+S\tau}\widehat{P}_{i+S\tau-1}\widetilde{\psi}_{i+S\tau}}, \ \widehat{\beta}_0 = 0, \\
\widehat{P}_{i+S\tau} &= \widehat{P}_{i+S\tau-1} - \frac{\widehat{P}_{i+S\tau-1}\widetilde{\psi}_{i+S\tau}\widetilde{\psi}'_{i+S\tau}\widehat{P}_{i+S\tau-1}}{\widehat{\sigma}_i^{2(i,\tau)} + \widetilde{\psi}'_{i+S\tau}\widehat{P}_{i+S\tau-1}\widetilde{\psi}_{i+S\tau}}, \ P_0 = K.I, \\
\widetilde{\psi}_{i+S\tau+1} &= \sum_{j=1}^{q} \mu_{i+1,j}\widetilde{\psi}_{i+S\tau+1-j} + \widehat{\varphi}_{i+S\tau+1}, \\
\widehat{\epsilon}_{i+S\tau} &= y_{i+S\tau} - \widehat{\beta}'_{i+S\tau}\widehat{\varphi}_{i+S\tau}, \\
& \text{Compute } \widehat{\varphi}_{i+S\tau} \text{ from (3.4.7).}
\end{aligned}$$

With $\widetilde{\psi}_t$ in place of $\widehat{\varphi}_t$, it is easy to see that the consistency of the resulting estimate provided by Algorithm 3.4.2 follows in the same manner as in the above, provided that

$$\sum_{k=1}^{t} \widetilde{a}_k \widetilde{b}_k \geq 0,$$

where a_t and b_t are defined similarly to (3.4.26) by

$$\widetilde{a}_t = -w^{-1}_{i_t,\tau_t}\widetilde{\psi}'_t\widetilde{\beta}_t, \tag{3.4.58a}$$

$$\widetilde{b}_t = w^{-1}_{i_t,\tau_t}\left(\widehat{\epsilon}_t - \varepsilon_t - \frac{1}{2}w_{i_t,\tau_t}\widetilde{a}_t\right). \tag{3.4.58b}$$

Using these notations, we can obtain similarly to Lemma 3.4.4 the following result, which gives the relation between $\widetilde{a}_t$ and $\widetilde{b}_t$.

Lemma 3.4.10

With $\widetilde{\psi}_t$ *defined by* (3.4.54) *and* $\widetilde{a}_t$ *and* $\widetilde{b}_t$ *given by* (3.4.58) *we have*

$$\widetilde{b}_t = \left(\frac{\mu_{i_t}(L)}{\theta_{i_t}(L)} - 1/2\right)\widetilde{a}_t, \tag{3.4.59}$$

where $\theta_i(L) = 1 - \sum_{j=1}^{p} \theta_{i,j}L^j$, $\mu_i(L) = 1 - \sum_{j=1}^{r} \mu_{i,j}L^j$ *and* $i_t \in \{1, ..., S\}$ *is such that* $t = i_t + S\tau_t$.

Proof

From (3.4.36) and (3.4.57) we have

$$\begin{aligned}(\widehat{\epsilon}_t - \varepsilon_t) &= -\theta_{i_t}(L)^{-1}\mu_t(L)\widetilde{\psi}_t'\widetilde{\beta}_t, \\ &= w_{i_t,\tau_t}\frac{\mu_{i_t}(L)}{\theta_{i_t}(L)}\widetilde{a}_t.\end{aligned} \tag{3.4.60}$$

Combining (3.4.60) and (3.4.58b), gives (3.4.59). ■

Thus we now establish without proof the consistency of the estimates obtained from Algorithm 3.4.2.

Theorem 3.4.3

Under the same assumptions of Theorem 3.4.1, *except that in B*2 *the transfer functions* $H_i(L) = \dfrac{1}{\theta_{i_t}(L)} - 1/2$ *may be replaced by* $\widetilde{H}_i(L) = \dfrac{\mu_i(L)}{\theta_i(L)} - 1/2$, *i.e.. that the following conditions*

$$\mathrm{Re}\left[\frac{\mu_s(e^{i\omega})}{\theta_s(e^{i\omega})} - \frac{1}{2}\right] > 0\,, \quad \forall\omega \in]-\pi, \pi[\,, \text{ for all } s = 1, ..., S \tag{3.4.61}$$

hold, the estimates obtained from Algorithm 3.4.2 *are strongly consistent.* ■

Next, we discuss the choice of the prefilter coefficient $\mu_{t,j}$. We know from Remark 3.4.5 that the positive realness condition (3.4.61) is equivalent to

$$\left|\frac{\theta_s(e^{i\omega})}{\mu_s(e^{i\omega})} - 1\right| < 1, \quad \forall\omega \in]-\pi, \pi[\,, \text{ for all } s = 1, ..., S, \tag{3.4.62}$$

which means that the transfer function $\dfrac{\theta_s(e^{i\omega})}{\mu_s(e^{i\omega})}$ is close to the unity. Hence a good choice of the prefilter $\mu_s(L)$ which can be considered to be a way of guaranteeing the positive realness of

$\widetilde{H}_s(L) = \dfrac{\mu_s(L)}{\theta_s(L)} - 1/2$ is to set

$$\mu_s(L) = \theta_s(L),\ s = 1, ..., S.$$

However as $\theta_i(L)$ is unknown, we can use

$$\mu_s(L) = \widetilde{\theta}_s(L)$$

for any preliminary estimated filter $\widetilde{\theta}_s(L)$ of $\theta_s(L)$. This approach has been suggested firstly by Ljung (1977a) and used by Friedlander (1982a), (1982b), Goodwin and Sin (1984) and Adams and Goodwin (1995).

3.4.4 Concluding remarks

In this section we have developped two recursive on-line estimation algorithm for $PARMA$ models, namely periodic extended least squares algorithms ($PELS$) without and with prefilter. The proposed algorithms, derived from the pseudo linear regression approach, generalize the well-known extended least squares algorithm (ELS) developed for stationary $ARMA$ models (cf, Panuska, 1968; Young, 1974) and coincide with the $PRLS$ algorithm (cf, section 3.3) in the case of PAR models. Similar algorithms developed for $PARMA$ models was already considered by Adams and Goodwin (1995). The main difference between the $PELS$ algorithms given here and those of Adams and Goodwin (1995) lies in the fact that our algorithms are based on the weighted least squares method instead of the ordinary least squares method, in order to take into account the heteroskedasticity of the innovations process. Moreover, our algorithms converge under positive real conditions which are weaker than the very strict passivity condition given in Adams and Goodwin (1995). An advantage of our algorithms is that they are computationally equivalent to the periodic recursive least squares algorithm ($PRLS$). Moreover, it has been shown that the $PELS$ estimators are strongly consistent under the positive realness condition on a certain transfer functions (see assumption $B2$, and equation (3.4.61)). The latter condition seems much simpler than the sufficient frequency domain condition for the very strict passivity assumption considered by Adams and Goodwin (1995). However the positive real condition $B2$ is too restrictive for higher order models, and the $PELS$ algorithm (Algorithm 3.4.1) will not guarantee the convergence for a certain choice of the moving average polynomials $\theta_i(L)$. The inclusion of a prefilter $\mu_t(L)$ in the $PELS$ algorithm can then relax the condition $B2$, while increasing the algorithmic complexity.

Another limitation of the pseudo linear regression approach when used in the presence of the moving average component, lies in the asymptotic inefficiency of the obtained estimates. Indeed, the limiting covariance matrix of a PLR algorithm is, to our knowledge, not fully understood at the present time (see also Ljung and Söderström, 1983 p. 231; Solo 1978, 1980). Moreover, Ljung and Söderström (1983, p. 231) showed via a particular counterexample, without giving the corresponding asymptotic distribution, that the PLR estimator (for an $MA(1)$ model) is asymptotically inefficient. To avoid these drawbacks, one can choose in the algorithm a particular prefilter, as outlined above, which on the one hand allows the consistency of the estimates without the need for the positive real condition, and on the other hand, makes these estimates asymptotically efficient. This leads to an algorithm which is an analogous to the well-known Recursive Maximum Likelihood (RML) method (cf, Söderström, 1973; Getler and Banyas, 1974) that we develop in the following section.

3.5 Periodic recursive maximum likelihood algorithm for $PARMA$ models

In the present section we shall develop a recursive on-line estimation algorithm for $PARMA$ models that avoids the two aforementioned drawbacks of the $PELS$ algorithm, namely the statistical inefficiency and the restrictive positive realness (and hence the strict passivity) conditions. Among the many existing approaches for building recursive algorithms, we chose the off-line approach for on-line methods, i.e. a family of recursive algorithms resulting from off-line estimation methods. The basic criterion for the algorithms of this class named "Recursive Predictor Error Method" ($RPEM$), is the minimization of the mean square error forecast, expressed in empirical form (see (3.2.4)). For many linear models such as the $ARMA$ and $PARMA$ models, the RPE criterion is the same as the conditional maximum likelihood criterion (when the innovation process is Gaussian), and gives estimators whose properties are equivalent to this one. This is the reason for which the obtained recursive algorithm is generally named "Recursive Maximum Likelihood" (RML). For other model structures such as conditionally heteroskedastic models, the RPE criterion does not coincide with the ML one, and the name RML will be inappropriate (see chapter 5). The RML algorithm is a recursive version of the very general class of maximum likelihood and prediction error methods. It was introduced by different authors using different approaches (cf, Söderström, 1973; Getler and Banyaz, 1974; Moore and Weiss, 1979) and was called Recursive Maximum Likelihood because of the similarity between the criterion functions, whereas the estimate obtained is not an exact maximum likelihood. Indeed, for the presence of a moving average component, the criterion function (3.3.2) would not be quadratic with respect to the parameters and the estimates cannot be obtained analytically as in the recursive least squares case (3.3.6). Instead, using a stochastic version of the Gauss-Newton optimization routine with some approximations, leads to the RML algorithm. As we can seen below, this algorithm can be seen as a generalization of the $PRLS$ algorithm when the underlying model includes a moving average component.

The remainder of this section is organized as follows, the first subsection recalls the principle of the class of recursive predictor error method, $RPEM$. In the second subsection, we build a recursive maximum likelihood algorithm type for $PARMA$ models, called periodic recursive maximum likelihood ($PRML$). Finally, the statistical properties of the $PRML$ estimates are studied in the third subsection.

3.5.1 Recursive prediction

We recall that the best prediction (in the mean square sense) of y_t, based on the past information until time $t-1$, is given by the conditional mean $E\left(y_t/\mathcal{F}_{t-1}\right)$, where $\mathcal{F}_t$ denotes the information until time t. For the $PARMA$ model

$$y_t - \sum_{j=1}^{p} \phi_{tj} y_{t-j} = \varepsilon_t - \sum_{j=1}^{q} \theta_{tj} \varepsilon_{t-j},\ t \in \mathbb{Z},$$

where $\{\varepsilon_t,\ t \in \mathbb{Z}\}$ is a periodic white noise with variance σ_t^2, this prediction is then given by

$$E\left(y_t/\mathcal{F}_{t-1}\right) = \sum_{j=1}^{p} \phi_{tj}^{*}\, y_{t-j} + \sum_{j=1}^{q} \theta_{tj}^{*}\, \varepsilon_{t-j},\ t \in \mathbb{Z}, \tag{3.5.1}$$

where β^* denotes the true value of the unknown parameter β. Since the innovations (ε_t) are unobservable and the data are generally available only for $t \geq 1$ the computation of $E\left(y_t/\mathcal{F}_{t-1}\right)$ is not possible via (3.5.1). However as in Section 2.3, the innovations $\{\varepsilon_t\}$ can be approximated, for a given β, by the *prediction error function* $\{\varepsilon_t(\beta)\}$ which satisfier the following stochastic difference equation

$$\begin{aligned}\varepsilon_t(\beta) &= y_t - \sum_{j=1}^{p} \phi_{tj} y_{t-j} - \sum_{j=1}^{q} \theta_{tj} \varepsilon_{t-j}(\beta), \ t \in \mathbb{Z}, \\ &= y_t - y_t(\beta),\end{aligned}$$

with the assumption that $\varepsilon_t(\beta^*) = \varepsilon_t$. The calculation of $\varepsilon_t(\beta)$ requires $p+q$ starting up values of the latter stochastic difference equation. For a causal model, the effect of these values decreases exponentially. Therefore they can be arbitrarily fixed without affecting the prediction error. Thus, we approximate again $\varepsilon_t(\beta)$ by the "*conditional prediction error function*" $\widehat{\varepsilon}_t(\beta)$ which satisfies the following "conditional" stochastic difference equation (a version of $PARMA$ model)

$$\widehat{\varepsilon}_t(\beta) = y_t - \sum_{j=1}^{p} \phi_{tj} y_{t-j} - \sum_{j=1}^{q} \theta_{tj} \widehat{\varepsilon}_{t-j}(\beta), \ t \geq 1 \tag{3.5.2}$$

with $\widehat{\varepsilon}_t(\beta) = y_t = 0$ for $t \leq 0$. Nevertheless, more sophisticated devices such as back-forecasting can be used to choose initial values. The *conditional prediction function* of y_t, denoted by $\widehat{y}_t(\beta)$, can then be deduced from (3.5.1) by

$$\widehat{y}_t(\beta) = \sum_{j=1}^{p} \phi_{tj}\, y_{t-j} - \sum_{j=1}^{q} \theta_{tj} \widehat{\varepsilon}_{t-j}(\beta), \ t \geq 1 \tag{3.5.3}$$

from which we have of course

$$\widehat{\varepsilon}_t(\beta) = y_t - \widehat{y}_t(\beta). \tag{3.5.4}$$

Combining (3.5.3) and (3.5.4), one can easily express the prediction $\widehat{y}_t(\beta)$ in the following recursive form :

$$\widehat{y}_t(\beta) - \sum_{j=1}^{q} \theta_{tj} \widehat{y}_{t-j}(\beta) = \sum_{i=1}^{r} \left(\phi_{tj} - \theta_{tj}\right) y_t, \ t \geq 1, \tag{3.5.5a}$$

with null initial values, where $r = \max(p,q)$. The same technique can be used for $\varepsilon_t(\beta)$ and $y_t(\beta)$ to lead to

$$y_t(\beta) - \sum_{j=1}^{q} \theta_{tj} y_{t-j}(\beta) = \sum_{i=1}^{r} \left(\phi_{tj} - \theta_{tj}\right) y_t, \ t \in \mathbb{Z}. \tag{3.5.5b}$$

The used criterion is as mentioned in Section 2.2 (see criterion (3.2.4)) the weighted least squares which we rewrite for convenience in an equivalent form

$$V_t(\beta) = \frac{1}{2} \sum_{k=1}^{t} \frac{\widehat{\varepsilon}_k^2(\beta)}{\sigma_k^2}. \tag{3.5.6}$$

Standard arguments (see Ljung and Söderström, 1983, p. 82) show that if the innovation process $\{\varepsilon_t,\ t \in \mathbb{Z}\}$ is Gaussian, the function $V_t(\beta)$ given by (3.5.6) is the negative log conditional likelihood function.

3.5.2 The $PRML$ algorithm

Though the derivation of an RML type algorithm is well known and can be found in the on-line estimation literature (see e.g. Söderström, 1973; Getler and Banyaz, 1974; Moore and Weiss, 1979; Ljung and Söderström, 1983), in order to have a self-contained work, we present a summary of our Periodic RML algorithm ($PRML$) and some of its properties. The derivation given here is similar to the RML version given by Ljung and Söderström (1983).

Let $\widehat{\beta}_{t-1}$ denote the recursive estimate of the $PARMA$ parameter β at time $t-1$, i.e. the one that (approximately) minimizes the function $V_{t-1}(\beta)$. We wish to obtain the estimate $\widehat{\beta}_t$ at time t, that (approximately) minimizes the criterion function $V_t(\beta)$. Developing $V_t(\beta)$ by means of the Taylor formula around $\widehat{\beta}_{t-1}$, we obtain

$$\begin{aligned} V_t(\beta) = V_t(\widehat{\beta}_{t-1}) &+ \frac{\partial V_t(\widehat{\beta}_{t-1})}{\partial \beta'} \left(\beta - \widehat{\beta}_{t-1}\right) + \tfrac{1}{2} \left(\beta - \widehat{\beta}_{t-1}\right)' \frac{\partial^2 V_t(\widehat{\beta}_{t-1})}{\partial \beta \partial \beta'} \left(\beta - \widehat{\beta}_{t-1}\right) \\ &+ o\left(\left|\beta - \widehat{\beta}_{t-1}\right|^2\right), \end{aligned} \tag{3.5.7}$$

where $o(x)$ denotes a function such that

$$\frac{o(x)}{|x|} \to 0, \text{ as } |x| \to 0.$$

Minimization of the right member of (3.5.7) with respect to β gives

$$\widehat{\beta}_t = \widehat{\beta}_{t-1} - \left[\frac{\partial^2 V_t(\widehat{\beta}_{t-1})}{\partial \beta \partial \beta'}\right]^{-1} \frac{\partial V_t(\widehat{\beta}_{t-1})}{\partial \beta} + o\left(\left|\widehat{\beta}_t - \widehat{\beta}_{t-1}\right|\right). \tag{3.5.8}$$

As in Section 2.3, let $\psi_t(\beta)$ and $\widehat{\psi}_t(\beta)$ denote, respectively, the negative derivative with respect to β, of the unconditional prediction error function $\varepsilon_t(\beta)$ and that of the conditional one $\widehat{\varepsilon}_t(\beta)$, that is

$$\begin{aligned} \psi_t(\beta) &= -\frac{\partial \varepsilon_t(\beta)}{\partial \beta} = \frac{\partial y_t(\beta)}{\partial \beta}, \\ \widehat{\psi}_t(\beta) &= -\frac{\partial \widehat{\varepsilon}_t(\beta)}{\partial \beta} = \frac{\partial \widehat{y}_t(\beta)}{\partial \beta}, \end{aligned}$$

and let also $R_t(\beta)$ be the second derivative of $V_t(\beta)$ with respect to β, i.e. $R_t(\beta) = \dfrac{\partial^2 V_t(\beta)}{\partial \beta \partial \beta'}$. Then from (3.5.6) we have

$$\begin{aligned} \frac{\partial V_t(\beta)}{\partial \beta} &= -\sum_{k=1}^{t} \frac{\widehat{\varepsilon}_t(\beta)}{\sigma_k^2} \widehat{\psi}_k(\beta), \\ &= \frac{\partial V_{t-1}(\beta)}{\partial \beta} - \frac{\widehat{\varepsilon}_t(\beta)}{\sigma_t^2} \widehat{\psi}_t(\beta), \end{aligned} \tag{3.5.9}$$

and by differentiating once more

$$R_t(\beta) = R_{t-1}(\beta) + \frac{1}{\sigma_t^2} \widehat{\psi}_t(\beta) \widehat{\psi}'_t(\beta) + \frac{\widehat{\varepsilon}_t(\beta)}{\sigma_t^2} \frac{\partial^2 \widehat{\varepsilon}_t(\beta)}{\partial \beta \partial \beta'}. \tag{3.5.10}$$

Therefore, equation (3.5.8) can be rewritten

$$\widehat{\beta}_t = \widehat{\beta}_{t-1} - \left[R_t\left(\widehat{\beta}_{t-1} \right) \right]^{-1} \frac{\partial V_t(\widehat{\beta}_{t-1})}{\partial \beta} + \left(o \left| \widehat{\beta}_t - \widehat{\beta}_{t-1} \right| \right).$$

This last equation, seeming to determine the vector of parameters in a recursive way, is not completely usable. We shall make some approximations to evaluate it.

i) For t sufficiently large we suppose that $\widehat{\beta}_t$ lies in a small neighborhood of $\widehat{\beta}_{t-1}$. On the basis of this assumption it seems reasonable to make the following approximations

$$\text{Neglect } o\left| \widehat{\beta}_t - \widehat{\beta}_{t-1} \right| \text{ in (3.5.8)}, \tag{3.5.11a}$$

and from the continuity of R_{t-1} we take

$$R_t\left(\widehat{\beta}_t \right) = R_t\left(\widehat{\beta}_{t-1} \right) \tag{3.5.11b}$$

ii) We assume that $\widehat{\beta}_{t-1}$ is the optimal estimate at time $t-1$, i.e.

$$R_{t-1}\left(\widehat{\beta}_{t-1}\right)=0. \tag{3.5.12}$$

iii) Finally we take

$$\frac{\widehat{\varepsilon}_t\left(\widehat{\beta}_{t-1}\right)}{\sigma_t^2}\frac{\partial^2\widehat{\varepsilon}_t\left(\widehat{\beta}_{t-1}\right)}{\partial\beta\partial\beta'}=0. \tag{3.5.13}$$

The motivation for using the approximation (3.5.13) lies in the fact that for the true value β^* we have

$$E\left(\frac{\partial^2\widehat{\varepsilon}_t\left(\beta^*\right)}{\partial\beta\partial\beta'}\widehat{\varepsilon}_t\left(\beta^*\right)\right)=0,$$

since $\widehat{\varepsilon}_t\left(\beta^*\right)\simeq\varepsilon_t$ (in the sense that $E\left(\widehat{\varepsilon}_t\left(\beta^*\right)-\varepsilon_t\right)^2\overset{t\to\infty}{\to}0$) and ε_t is uncorrelated with the past values of the process, in particular with $\dfrac{\partial^2\widehat{\varepsilon}_t\left(\beta^*\right)}{\partial\beta\partial\beta'}=-\dfrac{\partial^2\widehat{y}_t\left(\beta^*\right)}{\partial\beta\partial\beta'}$. Thus, when β is close to the true value β^*, the last term of (3.5.10) can be neglected since it makes an order of magnitude less contribution than the second term.

With the approximations (3.5.13) and (3.5.11b) inserted in (3.5.10) we can approximately evaluate $R_t\left(\widehat{\beta}_{t-1}\right)$. Let R_t be such an approximation. Then we have

$$R_t=R_{t-1}+\frac{1}{\sigma_t^2}\widehat{\psi}_t\left(\widehat{\beta}_{t-1}\right)\widehat{\psi}_t'\left(\widehat{\beta}_{t-1}\right). \tag{3.5.14}$$

Using now the assumption (3.5.12) into (3.5.9) we obtain

$$\frac{\partial V_t\left(\widehat{\beta}_{t-1}\right)}{\partial\beta}=-\frac{\widehat{\varepsilon}_t\left(\widehat{\beta}_{t-1}\right)}{\sigma_t^2}\widehat{\psi}_t\left(\widehat{\beta}_{t-1}\right). \tag{3.5.15}$$

Finally, with the approximations (3.5.11a) and (3.5.15) inserted into (3.5.8) we get the following system

$$\begin{aligned}\widehat{\beta}_t &= \widehat{\beta}_{t-1}+R_t^{-1}\widehat{\psi}_t\left(\widehat{\beta}_{t-1}\right)\frac{\widehat{\varepsilon}_t\left(\widehat{\beta}_{t-1}\right)}{\sigma_t^2}, \qquad (3.5.16)\\ R_t &= R_{t-1}+\frac{1}{\sigma_t^2}\widehat{\psi}_t\left(\widehat{\beta}_{t-1}\right)\widehat{\psi}_t'\left(\widehat{\beta}_{t-1}\right).\end{aligned}$$

It remains now to determine recursively $\widehat{\varepsilon}_t\left(\widehat{\beta}_{t-1}\right)$ and $\widehat{\psi}_t\left(\widehat{\beta}_{t-1}\right)$ in the sense of (3.2.1). For this, we again make some approximations. In light of (3.5.2) we approximate $\widehat{\varepsilon}_t\left(\widehat{\beta}_{t-1}\right)$ for a given $\widehat{\beta}_{t-1}$ by $\widehat{\varepsilon}_t$ obtained from the following stochastic difference equation (a modified version of (3.5.2))

$$\widehat{\varepsilon}_t = y_t - \sum_{j=1}^{p} \widehat{\phi}_{tj}^{(t-1)} y_{t-j} - \sum_{j=1}^{q} \widehat{\theta}_{tj}^{(t-1)} \widehat{\varepsilon}_{t-j},\ t \geq 1 \tag{3.5.17}$$

where the initial values are supposed to be known and where $\widehat{\phi}_{tj}^{(t-1)}$ and $\widehat{\theta}_{tj}^{(t-1)}$ are respectively the "autoregressive-part" and the "moving average-part" of the vector $\widehat{\beta}_{t-1}$.

To evaluate $\widehat{\psi}_t\left(\widehat{\beta}_{t-1}\right)$, deriving with respect to β both sides of the equality (3.5.5), we then obtain

$$\widehat{\psi}_t(\beta) - \sum_{j=1}^{q} \theta_{t,j} \widehat{\psi}_{t-j}(\beta) = \widehat{\varphi}_t(\beta), \tag{3.5.18a}$$

$$\psi_t(\beta) - \sum_{j=1}^{q} \theta_{t,j} \psi_{t-j}(\beta) = \varphi_t(\beta), \tag{3.5.18b}$$

where $\varphi_t(\beta)$ is similar to φ_t given by (3.4.2) except that the unmeasurable innovation process $\{\varepsilon_t\}$ is replaced by the prediction error function $\varepsilon_t(\beta)$ and so on. Equation (3.5.18a) represent a recursive filter with an infinite impulse response. This cannot be used by a recursive algorithm subject to the constraint (3.2.1). Instead of $\psi_t\left(\widehat{\beta}_{t-1}\right)$, we use an approximate $\overline{\psi}_t$, which is the solution to the following difference equation

$$\overline{\psi}_t - \sum_{j=1}^{q} \widehat{\theta}_{t,j}^{(t-1)} \overline{\psi}_{t-j} = \overline{\varphi}_t, \tag{3.5.19}$$

where $\overline{\varphi}_t$ is just φ_t given by (3.4.2) in which ε_t is replaced by $\widehat{\varepsilon}_t$ obtained from (3.5.17), i.e.

$$\overline{\varphi}_{i+S\tau} = \left(\underbrace{\mathbf{0}', ..., \mathbf{0}'}_{i-1 \text{ time}}, y_{i+S\tau-1}, ..., y_{i+S\tau-p}, -\widehat{\varepsilon}_{i+S\tau-1}, ..., -\widehat{\varepsilon}_{i+S\tau-q}, \underbrace{\mathbf{0}', ..., \mathbf{0}'}_{S-i \text{ time}}\right), \tag{3.5.20}$$

where $\mathbf{0}$, is the $(p+q)\times 1$ null vector, and where i and τ are such that $i = 1, 2, ..., S$ and $\tau \in \mathbb{Z}$. With the vector $\overline{\varphi}_t$ defined by (3.5.20), equation (3.5.17) can be rewritten

$$\widehat{\varepsilon}_t = y_t - \overline{\varphi}_t' \widehat{\beta}_{t-1}. \tag{3.5.21}$$

Thus, using $\widehat{\varepsilon}_t$ and $\overline{\psi}_t$ in the system (3.5.16) gives the following algorithm

$$\widehat{\varepsilon}_t = y_t - \overline{\varphi}_t' \widehat{\beta}_{t-1}, \tag{3.5.22a}$$

$$\widehat{\beta}_t = \widehat{\beta}_{t-1} + \overline{R}_t^{-1} \overline{\psi}_t \frac{\widehat{\varepsilon}_t}{\sigma_t^2}, \tag{3.5.22b}$$

$$\overline{R}_t = \overline{R}_{t-1} + \frac{1}{\sigma_t^2} \overline{\psi}_t \overline{\psi}_t', \tag{3.5.22c}$$

$$\overline{\psi}_{t+1} = \sum_{j=1}^{q} \widehat{\theta}_{t,j}^{(t-1)} \overline{\psi}_{t+1-j} + \overline{\varphi}_{t+1}, \tag{3.5.22d}$$

where $\overline{R}_t$ is obtained from R_t in (3.5.14) when replacing $\psi_t\left(\widehat{\beta}_{t-1}\right)$ by $\overline{\psi}_t$.

It is still possible to improve the recursions (3.5.22). In (3.5.22d) we use the "a priori" prediction error $\widehat{\varepsilon}_t = y_t - \overline{\varphi}_t' \widehat{\beta}_{t-1}$ in $\overline{\varphi}_t$ to compute $\overline{\psi}_t$. Since at this time we have already updated $\widehat{\beta}_t$, we can improve the estimate $\widehat{\varepsilon}_t$ by taking the new knowledge y_t into account. We then replace the estimated prediction error by the estimated residual $\widehat{\epsilon}_t$ which satisfies the difference equation (3.4.6), i.e.

$$\widehat{\epsilon}_t = y_t - \sum_{j=1}^{p} \widehat{\phi}_{tj}^{(t)} y_{t-j} - \sum_{j=1}^{q} \widehat{\theta}_{tj}^{(t)} \widehat{\epsilon}_{t-j}, \ t \geq 1, \tag{3.5.23a}$$

with null initial values. Now rewrite (3.5.23a) in a vectorial form

$$\widehat{\epsilon}_t = y_t - \widehat{\varphi}_t' \widehat{\beta}_t, \tag{2.5.23b}$$

where $\widehat{\varphi}_t$ is the one defined in (3.4.7). Let $\widehat{\psi}_t$ be the vector obtained from $\overline{\psi}_t$ by filtering the vector $\widehat{\varphi}_t$ instead of $\overline{\varphi}_t$, through the current estimate of the θ_t-polynomial, i.e.

$$\widehat{\psi}_t = \sum_{j=1}^{q} \widehat{\theta}_{t,j}^{(t-1)} \widehat{\psi}_{t-j} + \widehat{\varphi}_t \tag{3.5.24}$$

With these improved estimates of $\widehat{\psi}_t$ and $\widehat{\epsilon}_t$, $\overline{R}_t$ given by (3.5.22c) is now replaced by $\widehat{R}_t$ with

$$\widehat{R}_t = \widehat{R}_{t-1} + \frac{1}{\sigma_t^2} \widehat{\psi}_t \widehat{\psi}_t'$$

Before stating our modified algorithm of (3.5.22), it remains to estimate the unknown variances σ_t^2 as in the above by means of the recursive relation (3.3.10) (cf, Section 3.3). Thus we obtain the following recursions

$$\widehat{\varepsilon}_t = y_t - \widehat{\varphi}_t' \widehat{\beta}_{t-1}, \tag{3.5.25a}$$

$$\widehat{\sigma}_i^{2(i,\tau)} = \begin{cases} \dfrac{\tau}{\tau+1}\,\widehat{\sigma}_i^{2(i,\tau-1)} + \dfrac{1}{\tau+1}\widehat{\varepsilon}_{i+S\tau}^2, \ 0 < \tau \le \delta, \\ \dfrac{\tau-\delta}{\tau+1-\delta}\,\widehat{\sigma}_i^{2(i,\tau-1)} + \dfrac{1}{\tau+1-\delta}\widehat{\varepsilon}_{i+S\tau}^2, \ \tau > \delta, \end{cases} \tag{3.5.25b}$$

$$\widehat{\beta}_t = \widehat{\beta}_{t-1} + \widehat{R}_t^{-1}\widehat{\psi}_t \frac{\widehat{\varepsilon}_t}{\widehat{\sigma}_{i_t}^{2(i_t,\tau_t)}}, \tag{3.5.25c}$$

$$\widehat{R}_t = \widehat{R}_{t-1} + \frac{1}{\widehat{\sigma}_{i_t}^{2(i_t,\tau_t)}}\widehat{\psi}_t\widehat{\psi}_t', \tag{3.5.25d}$$

$$\widehat{\epsilon}_t = y_t - \widehat{\varphi}_t'\widehat{\beta}_t, \tag{3.5.25e}$$

$$\widehat{\psi}_{t+1} = \sum_{j=1}^{q} \widehat{\theta}_{t+1,j}^{(t)} \widehat{\psi}_{t+1-j} + \widehat{\varphi}_{t+1}. \tag{3.5.25f}$$

Just as in Sections 3.3 and 3.4, we can of course introduce $\widehat{P}_t = \widehat{R}_t^{-1}$ and update $\widehat{P}_t$ using the MIL (see Appendix A). We then obtain the following RML type algorithm which we call Periodic RML, ($PRML$).

Algorithm 3.5.1 (The $PRML$ algorithm)

The periodic recursive maximum likelihood ($PRML$) algorithm for $PARMA$ models is given by the following recursive relations

$$\begin{cases} \widehat{\varepsilon}_{i+S\tau} = y_{i+S\tau} - \widehat{\beta}_{i+S\tau-1}'\widehat{\varphi}_{i+S\tau}, \\ \widehat{\sigma}_i^{2(i,\tau)} = \begin{cases} \dfrac{\tau}{\tau+1}\,\widehat{\sigma}_i^{2(i,\tau-1)} + \dfrac{1}{\tau+1}\widehat{\varepsilon}_{i+S\tau}^2, \ 0 < \tau \le \delta, \\ \dfrac{\tau-\delta}{\tau+1-\delta}\,\widehat{\sigma}_i^{2(i,\tau-1)} + \dfrac{1}{\tau+1-\delta}\widehat{\varepsilon}_{i+S\tau}^2, \ \tau > \delta, \end{cases} \\ \widehat{\beta}_{i+S\tau} = \widehat{\beta}_{i+S\tau-1} + \dfrac{\widehat{P}_{i+S\tau}\widehat{\psi}_{i+S\tau}\widehat{\varepsilon}_{i+S\tau}}{\widehat{\sigma}_i^{2(i,\tau)} + \widehat{\psi}_{i+S\tau}'\widehat{P}_{i+S\tau-1}\widehat{\psi}_{i+S\tau}}, \ \widehat{\beta}_0 = 0, \\ \widehat{P}_{i+S\tau} = \widehat{P}_{i+S\tau-1} - \dfrac{\widehat{P}_{i+S\tau-1}\widehat{\psi}_{i+S\tau}\widehat{\psi}_{i+S\tau}'\widehat{P}_{i+S\tau-1}}{\widehat{\sigma}_i^{2(i,\tau)} + \widehat{\psi}_{i+S\tau}'\widehat{P}_{i+S\tau-1}\widehat{\psi}_{i+S\tau}}, \ \widehat{P}_0 = K.I, \\ \widehat{\psi}_{t+1} = \sum_{j=1}^{q} \widehat{\theta}_{t+1,j}^{(t)}\widehat{\psi}_{t+1-j} + \widehat{\varphi}_{t+1}, \ (t = i + S\tau), \\ \widehat{\epsilon}_{i+S\tau} = y_{i+S\tau} - \widehat{\beta}_{i+S\tau}'\widehat{\varphi}_{i+S\tau}, \\ \text{Compute } \widehat{\varphi}_{i+S\tau} \text{ from (3.4.7),} \end{cases} \tag{3.5.26}$$

where $\delta = (p+q)\,S$, K *is a large enough number and* $\widehat{\sigma}_i^{2(i,0)} = \widehat{\varepsilon}_i^2$.

Remark 3.5.1

It is of interest to compare the $PRML$ algorithm with the $PELS$ algorithms (Algorithm 3.4.1 and Algorithm 3.5.2).

i) The only difference between the $PRML$ algorithm (Algorithm 3.5.1) and the $PELS$ one (Algorithm 3.4.1) is that the vector $\widehat{\varphi}_t$ in Algorithm 3.4.1 is replaced by $\widehat{\psi}_t$ (in the updating equation of $\widehat{\beta}_t$ and P_t) by filtering it through the current estimate of the θ_t-polynomial in (3.5.24), that is

$$\widehat{\theta}_t^{(t-1)}(L)\widehat{\psi}_t = \widehat{\varphi}_t. \tag{3.5.27}$$

ii) Similarly, Algorithm 3.5.1 can be obtained from Algorithm 3.4.2 by setting

$$\mu_t(L) = \widehat{\theta}_t^{(t-1)}(L), \tag{3.5.28}$$

instead of

$$\mu_t(L) = \widetilde{\theta}_t(L).$$

In the following subsections we investigate what this difference means for the convergence properties of the respective algorithms.

3.5.3 Large sample properties of the $PRML$ estimates

Consistency of the $PRML$ estimator

It is clear that the consistency of the $PRML$ estimator follows immediately from that of the $PRLS$, because of the link between these algorithms as pointed out in Remark 3.5.1. Thus the same conditions of Theorem 3.4.1 will be considered except that the condition $B2$ is replaced by

$$\mathrm{Re}\left[\frac{\widehat{\theta}_t^{(t-1)}(e^{i\omega})}{\theta_s(e^{i\omega})} - \frac{1}{2}\right] > 0\ ,\quad \forall\omega \in \,]-\pi, \pi[\ ,\ \text{ for all } s = 1, ..., S, \tag{3.5.29}$$

which holds provided that the filter $\widehat{\theta}_t^{(t-1)}(L)$ is stable. However, the $PRML$ algorithm as it stands, does not guarantee the stability of $\widehat{\theta}_t^{(t-1)}(L)$ and hence we must prevent the recursive estimate $\widehat{\beta}_t$ from getting outside the stability region.

Thus, for the $PRML$ algorithm we assume (unlike the $PELS$) that $B4$ holds even if we replace $\widehat{\sigma}_{i_t}^{-2(i_t,\tau_t)}$ by $\widehat{\sigma}_{i_t}^{-2(i_t,\tau_t-1)}$.

Theorem 3.5.1

Under assumptions $B1$, $B3$ and $B4$ we have

$$t^{1/2-\delta}\left(\widehat{\beta}_t - \beta^*\right) \stackrel{t\to\infty}{\to} 0, \ a.s. \tag{3.5.30}$$

for any $\delta > 0$.

Proof

The proof follows from similar arguments used in the proofs of Theorem 3.4.1 and Theorem 3.4.2. ■

Once the consistency of the $PRML$ estimate was established under the aforementioned assumptions, a natural question is to determine the "speed of convergence" to the limit. This is usually expressed in terms of an asymptotic distribution for the estimate. Next, we deal with this important point.

Asymptotic efficiency of the $PRML$ estimator

As is well known in the literature of the $ARMA$ recursive estimation, the ELS has an asymptotic error variance strictly larger than the Cramér-Rao bound (cf, Solo, 1978; Ljung and Söderström, 1983). It is also known that the RML type algorithm provides asymptotically efficient estimates, i.e. the covariance of the estimate converges to the Cramér-Rao lower bound (cf, Solo, 1981, Söderström et al., 1978, Ljung and Söderström, 1983). For $ARMAX$ models, Solo (1981) established the asymptotic efficiency of the monitored RML recursion under the assumption of the existence of higher order moments (strictly higher than the fourth moment). His proof was based on the introduction of a fictitious term (related to the recursion) which obeys a central limit theorem (CLT). The term is subtracted from the normed up recursion and the remainder shown to converge to zero (a.s.). Ljung and Söderström (1983) have used this technique to prove the asymptotic efficiency of the more general recursive predictor error algorithm (RPE), which is devoted to a general class of time-invariant linear models. However, to simplify their proof, they considered the very restrictive assumption that the underlying processes are bounded a.s. Such an assumption will exclude a large class of stochastic processes such as the Gaussian one. In our $PARMA$ case, we adopt Solo (1981)'s technique to show the asymptotic efficiency of the $PRML$ estimator. The adaptation is relatively straightforward but the details are tedious. Certain steps of our proof are nearly identical to the proofs stated in Solo (1981) for other algorithms and can then be omitted.

However, to have a self-contained demonstration, we prefer to expose the entire proof. Consider the following assumptions on the $PARMA$ model.

$B5$: There exists a small $0 < \delta^* < 1/3$, such that

$$\sup_{t\in\mathbb{Z}} E\left(|\varepsilon_t|^{4/1-3\delta^*}\right) < K < \infty. \tag{3.5.31}$$

$B6$: The $PARMA$ model is not redundant in the sense that the polynomials $\phi_s(L)$ and $\theta_s(L)$ have no common roots for any $s \in \{1, ..., S\}$.

Notice that when the process $\{\varepsilon_t,\ t \in \mathbb{Z}\}$ is periodically distributed, the moment assumption (3.5.31) can be replaced by

$$\sup_{s\in\{1,\dots,S\}} E\left(|\varepsilon_s|^{4/1-3\delta^*}\right) < K < \infty.$$

We also note that condition $B6$ is made to ensure the existence of the asymptotic covariance matrix of the $PRML$ estimator, which is related to the inverse of the limiting Fisher information matrix (See Theorem 1.4.1). Using the previous assumptions and notations, we can state the following result which establishes the asymptotic normality and asymptotic efficiency of the $PRML$ estimator.

Theorem 3.5.2

Consider the $PRML$ algorithm (3.5.26) subject to $B1$ and $B3 - B6$. Then

$$\sqrt{\frac{t}{S}}\left(\widehat{\beta}_t - \beta^*\right) \underset{t\longrightarrow\infty}{\xrightarrow{D}} \mathcal{N}\left(0, \left[F\left(\beta^*, \boldsymbol{\sigma}^*\right)\right]^{-1}\right), \tag{3.5.32}$$

where

$$\begin{aligned} F(\beta, \boldsymbol{\sigma}) &= \sum_{i=1}^{S} \frac{1}{\sigma_i^2} E\left[\frac{\partial\varepsilon_{i+S\tau}(\beta)}{\partial\beta}\frac{\partial\varepsilon_{i+S\tau}(\beta)}{\partial\beta'}\right], \\ &= \sum_{i=1}^{S} \frac{1}{\sigma_i^2} E\left[\psi_{i+S\tau}(\beta)\,\psi'_{i+S\tau}(\beta)\right], \end{aligned}$$

denotes the asymptotic Fisher information matrix for a $PARMA$ model. ■

Before giving the proof of the latter theorem, we need to show several lemmas.

Substract β^* from $(3.5.25c)$ and multiply the resulting equation by $\widehat{R}_t$ to obtain

$$\begin{aligned}\widehat{R}_t\left(\widehat{\beta}_t-\beta^*\right) &= \widehat{R}_t\left(\widehat{\beta}_{t-1}-\beta^*\right)+\widehat{\psi}_t\frac{\widehat{\varepsilon}_t}{\widehat{\sigma}_{i_t}^{2(i_t,\tau_t)}},\\ &= \widehat{R}_{t-1}\left(\widehat{\beta}_{t-1}-\beta^*\right)+\frac{1}{\widehat{\sigma}_{i_t}^{2(i_t,\tau_t)}}\widehat{\psi}_t\widehat{\psi}_t'\left(\widehat{\beta}_{t-1}-\beta^*\right)+\widehat{\psi}_t\frac{\widehat{\varepsilon}_t}{\widehat{\sigma}_{i_t}^{2(i_t,\tau_t)}}.\end{aligned}$$

The latter expression can be summed from $t=0$ to t, giving

$$\begin{aligned}\widehat{R}_t\widetilde{\beta}_t &= \widehat{R}_0\widetilde{\beta}_0+\sum_{k=1}^{t}\widehat{\psi}_k\frac{\left(\widehat{\psi}_k'\widetilde{\beta}_{k-1}+\widehat{\varepsilon}_k\right)}{\widehat{\sigma}_{i_k}^{2(i_k,\tau_k)}},\\ &= \widehat{R}_0\widetilde{\beta}_0+\sum_{k=1}^{t}\widehat{\psi}_k\frac{\left(\widehat{\psi}_k'\widetilde{\beta}_{k-1}+\widehat{\varepsilon}_k-\varepsilon_k\right)}{\widehat{\sigma}_{i_k}^{2(i_k,\tau_k)}}+\sum_{k=1}^{t}\widehat{\psi}_k\frac{\varepsilon_k}{\widehat{\sigma}_{i_k}^{2(i_k,\tau_k)}},\end{aligned}$$

where the notation $\widetilde{\beta}_t=\widehat{\beta}_t-\beta^*$ has been used. Rearranging this last equality leads to

$$\begin{aligned}t^{1/2}\widetilde{\beta}_t = t^{-1/2}\left(\frac{1}{t}\widehat{R}_t\right)^{-1}\widehat{R}_0\widetilde{\beta}_0+t^{-1/2}\left(\frac{1}{t}\widehat{R}_t\right)^{-1}\sum_{k=1}^{t}\widehat{\psi}_k\frac{\varepsilon_k}{\widehat{\sigma}_{i_k}^{2(i_k,\tau_k)}}+\\ t^{-1/2}\left(\frac{1}{t}\widehat{R}_t\right)^{-1}\sum_{k=1}^{t}\widehat{\psi}_k\frac{\left(\widehat{\psi}_k'\widetilde{\beta}_{k-1}+\widehat{\varepsilon}_k-\varepsilon_k\right)}{\widehat{\sigma}_{i_k}^{2(i_k,\tau_k)}}.\end{aligned} \tag{3.5.33}$$

Now, the following lemmas are devoted to analyzing each term of (3.5.33). Lemma 3.5.1 shows that under the assumptions $B4$ and $B5$, the norms $\|\widehat{\varphi}_k\|$ and $\left\|\widehat{\psi}_k\right\|$ verify a moment condition similar to $B5$. With these results, Lemma 3.5.2 establishes the asymptotic nullity of the difference between the gradient of the unconditional prediction error $\psi_t(\beta^*)$ and its approximate $\widehat{\psi}_t$. Similarly, Lemma 3.5.3 gives a similar result between the Hessian $\widehat{R}_t(\beta^*)$ and its approximate $\widehat{R}_t$. Based on this result, the limit of $\frac{1}{t}\widehat{R}_t$ is shown to be equal to the asymptotic Fisher information matrix of a $PARMA$ model. Finally, Lemma 3.5.4 shows that the last term of the right member of (3.5.33) is asymptotically negligible $(a.s.)$, while Lemma 3.5.5 gives a central limit theorem for the second term of the right member of (3.5.33). In the following proofs, C, λ and D will denote any positive constant not necessarily the same when appearing in different terms.

Lemma 3.5.1

For the $PRML$ algorithm (3.5.26) *suppose that $B4$ and $B5$ hold and that the $PARMA$ model is stable (causal and invertible). Then*

$$i) \qquad \sup_{t\in\mathbb{Z}} E\left(\|\widehat{\varphi}_t\|^{4/1-3\delta^*}\right) < \infty, \tag{3.5.34a}$$

$$ii) \qquad \sup_{t\in\mathbb{Z}} E\left(|\widehat{\epsilon}_t - \varepsilon_t|^{4/1-3\delta^*}\right) < \infty, \tag{3.5.34b}$$

$$iii) \qquad \sup_{t\in\mathbb{Z}} E\left(\left\|\widehat{\psi}_t\right\|^{4/1-3\delta^*}\right) < \infty. \tag{3.5.34c}$$

Proof

To see (3.5.34a), it suffices to show that

$$\sup_{t\in\mathbb{Z}} E\left(|y_t|^{4/1-3\delta^*}\right) < \infty, \tag{3.5.35a}$$

$$\sup_{t\in\mathbb{Z}} E\left(|\widehat{\epsilon}_t|^{4/1-3\delta^*}\right) < \infty, \tag{3.5.35b}$$

and then i) follows from the C_r-inequality (see e.g. Appendix B, and Sen and Singer, 1993, p. 21).

Now since the $PARMA$ model

$$y_t = \phi_t^{-1}(L)\,\theta_t(L)\,\varepsilon_t,$$

is supposed causal, there must exist S-periodic constants (λ_t) and (D_t) satisfying $\prod_{i=1}^{S}|\lambda_i| < 1$, such that

$$\begin{aligned} y_t &= \sum_{j=0}^{t-1}\left(\prod_{i=0}^{j-1}\lambda_{t-i}\right)\varepsilon_{t-j} + \left(\prod_{i=0}^{t-1}\lambda_{t-i}\right)D_t, \\ &= \sum_{j=0}^{t-1}\left(\prod_{l=1}^{S}\lambda_l\right)^{\tau^{(j)}}\left(\prod_{l=1}^{n^{(j)}}\lambda_l\right)\varepsilon_{t-j} + \left(\left(\prod_{l=1}^{S}\lambda_l\right)^{\tau^{(t)}}\prod_{l=1}^{n^{(t)}}\lambda_l\right)D_t, \end{aligned} \tag{3.5.36}$$

where the integers $n^{(j)}$, $n^{(t)}$, $\in \{1, ..., S\}$ and $\tau^{(j)}$, $\tau^{(t)}$ are such that $t = n^{(t)} + S\tau^{(t)}$ and $j =$

$n^{(j)} + S\tau^{(j)}$. Note that the term $\left(\prod_{i=0}^{t-1} \lambda_{t-i}\right) D_t$ in (3.5.36) accounts for initial conditions. Now

$$\begin{aligned}
|y_t| &\leq \sum_{j=0}^{t-1} \left(\prod_{l=1}^{S} |\lambda_l|\right)^{[j/S]} \left(\prod_{l=1}^{n^{(j)}} |\lambda_l|\right) |\varepsilon_{t-j}| + \left(\left(\prod_{l=1}^{S} |\lambda_l|\right)^{[t/S]} \prod_{l=1}^{n^{(t)}} |\lambda_l|\right) |D_t|, \\
&\leq C \sum_{j=1}^{t-1} \left(\prod_{l=1}^{S} |\lambda_l|\right)^{[j/S]} |\varepsilon_{t-j}| + C \left(\prod_{l=1}^{S} |\lambda_l|\right)^{[t/S]} |D_t|, \\
&\leq C \sum_{j=1}^{t-1} \lambda^{[j/S]} |\varepsilon_{t-j}| + D\lambda^{[t/S]}, \\
&\leq C \sum_{j=0}^{t-1} \lambda^{j} |\varepsilon_{t-j}| + D\lambda^{t}, \qquad (3.5.37) \\
&= C \sum_{j=1}^{t} \lambda^{t-j} |\varepsilon_{t}| + D\lambda^{t}.
\end{aligned}$$

Using Hölder's inequality in the right member of relation (3.5.37) (cf, Appendix B) leads to

$$\sum_{j=0}^{t-1} \left(\lambda^{j/2} |\varepsilon_{t-j}|\right) \lambda^{j/2} + D\lambda^t \leq \left(\sum_{j=0}^{t-1} \left(|\varepsilon_{t-j}| \lambda^{j/2}\right)^{4/1-3\delta^*}\right)^{\frac{1-3\delta^*}{4}} \left(\sum_{j=0}^{t-1} \left(\lambda^{j/2}\right)^{\frac{4}{3+3\delta^*}}\right)^{\frac{3+3\delta^*}{4}} + D\lambda^t.$$

Now from (3.5.37) with the C_r-inequality, we obtain

$$|y_t|^{\frac{4}{1-3\delta^*}} \leq 2^{\frac{4}{1-3\delta^*}} \left[\left(\sum_{j=0}^{t-1} \left(|\varepsilon_{t-j}| \lambda^{j/2}\right)^{4/1-3\delta^*}\right) \left(\sum_{j=0}^{t-1} \left(\lambda^{j/2}\right)^{\frac{4}{3+3\delta^*}}\right) + D\lambda^{\frac{4t}{1-3\delta^*}}\right],$$

which implies (3.5.35a).

Turn now to (3.5.35b). Rewrite (3.5.23a) as follows

$$\widehat{\theta}_t^{(t)}(L)\,\widehat{\epsilon}_t = \widehat{\phi}_t^{(t)}(L)\, y_t,$$

since by $B4$, $\widehat{\beta}_t$ lies in the stability region then as for (3.5.26), there must exist a positive constant $\lambda < 1$ and another D such that

$$|\widehat{\epsilon}_t| \leq C \sum_{j=1}^{t} \lambda^{t-j} |y_t| + D\lambda^t, \qquad (3.5.38)$$

whereupon (using a similar argument as in proving (3.5.35a)) (3.5.34a), Hölder's inequality and the C_r-inequality will ensure that (3.5.34a) holds. This completes the proof of i).

ii) Relation (3.5.34b) is a direct consequence of assumption $B5$, relation (3.5.34a) and the C_r-inequality.

iii) From (3.5.27), it can be seen that $\widehat{\psi}_t$ is obtained from $\widehat{\varphi}_t$ by filtering this last through the filter $\widehat{\theta}_t^{(t-1)}(L)$, which is supposed stable, in view of $B5$. Take one component of $\widehat{\psi}_t$ say $\overline{\epsilon}_t$ with

$$\widehat{\theta}_t^{(t-1)}(L)\,\overline{\epsilon}_t = \widehat{\epsilon}_t,$$

then for some positive $\lambda < 1$ we have

$$|\overline{\epsilon}_t| \leq C\sum_{j=0}^{t-1}\lambda^j\,|y_{t-j}| + D\lambda^t, \tag{3.5.39}$$

and by the same argument that shows (3.5.38) implies (3.5.34a) we see that

$$\sup_{t\in\mathbb{Z}} E\left(|\overline{\epsilon}_t|^{4/1-3\delta^*}\right) < \infty,$$

follows from (3.5.39). Similar arguments follow for the other component of $\widehat{\psi}_t$, that is

$$\sup_{t\in\mathbb{Z}} E\left(|\overline{y}_t|^{4/1-3\delta^*}\right) < \infty,$$

where $\overline{y}_t$ is such that

$$\widehat{\theta}_t^{(t-1)}(L)\,\overline{y}_t = y_t.$$

Using again the C_r-inequality will lead to iii). ■

Lemma 3.5.2

Under the same assumptions of Lemma 3.5.1 we have

$$i) \qquad t^4\,|\widehat{\epsilon}_t - \varepsilon_t| \overset{t\to\infty}{\to} 0 \quad a.s., \tag{3.5.40a}$$

$$ii) \qquad \psi_t(\beta^*) - \widehat{\psi}_t \overset{t\to\infty}{\to} 0 \quad a.s., \tag{3.5.40b}$$

where $\psi_t(\beta^*)$ *is given by* (3.5.18b).

Proof

i) Rewrite equations (3.4.1) for the true value of the parameter β^* and (3.5.23a) as follows

$$\begin{aligned}\theta_t^*(L)\,\varepsilon_t &= \phi_t^*(L)\,y_t,\\ \widehat{\theta}_t^{(t)}(L)\,\widehat{\epsilon}_t &= \widehat{\phi}_t^{(t)}(L)\,y_t,\end{aligned}$$

Thus we obtain

$$\begin{aligned}\theta_t^*(L)\left(\widehat{\epsilon}_t-\varepsilon_t\right) &= \left(\widehat{\phi}_t^{(t)}(L)-\phi_t^*(L)\right)y_t-\left(\widehat{\theta}_t^{(t)}(L)-\theta_t^*(L)\right)\widehat{\epsilon}_t,\\ &= \widehat{\varphi}_t'\left(\widehat{\beta}_t-\beta^*\right). \qquad (3.5.41a)\end{aligned}$$

Since the $PARMA$ model is assumed invertible, then similarly to (3.5.38), we must have for some $0<\lambda<1$ and another D

$$|\widehat{\epsilon}_t-\varepsilon_t|\leq\sum_{j=1}^{t}\lambda^{t-j}\left\|\widehat{\varphi}_j\right\|\left\|\widehat{\beta}_j-\beta^*\right\|+D\lambda^t, \qquad (3.5.41b)$$

using the result of Theorem 3.5.1 for $\delta=\dfrac{\delta^*}{2}$

$$\left\|\widehat{\beta}_t-\beta^*\right\|=t^{-1/2+\delta^*/2}o_t(1), \qquad (3.5.42)$$

for some negligible positive function $o_t(1)$, we can write (3.5.41b)

$$\begin{aligned}t^4|\widehat{\epsilon}_t-\varepsilon_t| &\leq t^4\sum_{j=1}^{t}\lambda^{t-j}\left\|\widehat{\varphi}_j\right\|j^{-1/2+\delta^*/2}o_j(1)+D\lambda^t,\\ &= \sum_{j=1}^{t}\left(\left(\frac{t}{j}\right)^4\lambda^{t-j}o_j(1)\right)\left(j^{-1/4+\delta^*/2}\left\|\widehat{\varphi}_j\right\|\right)+D\lambda^t.\end{aligned}$$

Set $a_{tj}=\left(\frac{t}{j}\right)^4\lambda^{t-j}o_j(1)$. It is clear that $a_{tj}\overset{t\to\infty}{\to}0$ for every fixed $1\leq j\leq t$ and that $\sum_{j=1}^{t}\left(\frac{t}{j}\right)^4\lambda^{t-j}o_j(1)$ is bounded for all t. Thus if we can show that

$$t^{-1/4+\delta^*/2}\left\|\widehat{\varphi}_t\right\|\overset{t\to\infty}{\to}0\ a.s., \qquad (3.5.43)$$

then by the Toeplitz lemma (see Appendix B, or e.g. Hall and Heyde, 1980, p. 31) we obtain the desired result (3.5.40a). Now, we will show that this last fact follows from result i) of Lemma 3.5.1 (see (3.5.34a)). Indeed, by the Markov inequality (cf, Appendix B) we have from (3.5.40a)

$$\begin{aligned}\text{For every positive } \eta,\ P\left(\|\widehat{\varphi}_t\| > t^{\frac{1}{4}-\frac{\delta^*}{2}}\eta\right) &\leq \frac{E\left(\|\widehat{\varphi}_t\|^{\frac{4}{1-3\delta^*}}\right)}{t^{\frac{1-2\delta^*}{1-3\delta^*}}\eta^{\frac{4}{1-3\delta^*}}}, \\ &= \frac{1}{t^{1+\frac{\delta^*}{1-3\delta^*}}}\frac{E\left(\|\widehat{\varphi}_t\|^{\frac{4}{1-3\delta^*}}\right)}{\eta^{\frac{4}{1-3\delta^*}}},\end{aligned}$$

using now (3.5.40a) and the fact that $\sum_{t\geq 1} t^{-1-\frac{\delta^*}{1-3\delta^*}} < \infty$ we then obtain

$$\sum_{t\geq 1} P\left(t^{\frac{1}{4}-\frac{\delta^*}{2}}\|\widehat{\varphi}_t\| > \eta\right) < \infty,$$

and by the Borel-Cantelli lemma (see appendix B, or Theorem 2.1.1 of Stout 1974) we conclude that (3.5.43) is true. Thus part i) of the lemma is established.

ii) Recall (3.5.25) and (3.5.18b) for $\beta = \beta^*$

$$\begin{aligned}\widehat{\theta}_t^{(t-1)}(L)\widehat{\psi}_t &= \widehat{\varphi}_t, \\ \theta_t^*(L)\,\psi_t(\beta^*) &= \varphi_t(\beta^*),\end{aligned}$$

with

$$\varphi_{i+S\tau}(\beta) = \Bigg(\underbrace{\mathbf{0}',...,\mathbf{0}'}_{i-1\text{ time}}, y_{i+S\tau-1},...,y_{i+S\tau-p}, -\varepsilon_{i+S\tau-1}(\beta),...,-\varepsilon_{i+S\tau-q}(\beta), \underbrace{\mathbf{0}',...,\mathbf{0}'}_{S-i\text{ time}}\Bigg)',$$

where $\mathbf{0}$ is the $(p+q)\times 1$ null vector and i and τ are such that $i = 1, 2, ..., S$ and $\tau \in \mathbb{Z}$. Notice that since we assumed $\varepsilon_t(\beta^*) = \varepsilon_t$ then $\varphi_t(\beta^*) = \varphi_t$, where φ_t is given by (3.4.2). The vectors $\psi_t(\beta^*)$ and $\widehat{\psi}_t$ have the form

$$\begin{aligned}\psi_{i+S\tau}(\beta^*) &= \Bigg(\underbrace{\mathbf{0}',...,\mathbf{0}'}_{i-1\text{ time}}, y^*_{i+S\tau-1},...,y^*_{i+S\tau-p}, -e^*_{i+S\tau-1},...,-e^*_{i+S\tau-q}, \underbrace{\mathbf{0}',...,\mathbf{0}'}_{S-i\text{ time}}\Bigg), \\ \widehat{\psi}_{i+S\tau} &= \Bigg(\underbrace{\mathbf{0}',...,\mathbf{0}'}_{i-1\text{ time}}, \widehat{y}_{i+S\tau-1},...,\widehat{y}_{i+S\tau-p}, -\widehat{e}_{i+S\tau-1},...,-\widehat{e}_{i+S\tau-q}, \underbrace{\mathbf{0}',...,\mathbf{0}'}_{S-i\text{ time}}\Bigg),\end{aligned}$$

where

$$\theta_t^*(L)\, y_t^* = y_t, \tag{3.5.44a}$$

$$\theta_t^*(L)\, e_t^* = \varepsilon_t, \tag{3.5.44b}$$

$$\widehat{\theta}_t^{(t-1)}(L)\, \widehat{y}_t = y_t, \tag{3.5.44c}$$

$$\widehat{\theta}_t^{(t-1)}(L)\, \widehat{e}_t = \widehat{\epsilon}_t. \tag{3.5.44d}$$

Hence the components of $\psi_t(\beta^*) - \widehat{\psi}_t$ are of three types : either 0, $y_t^* - \widehat{y}_t$, or $e_t^* - \widehat{e}_t$. To establish (3.5.40b) it suffices to show

$$e_t^* - \widehat{e}_t \overset{t\to\infty}{\to} 0 \;\; a.s., \tag{3.5.45a}$$

$$y_t^* - \widehat{y}_t \overset{t\to\infty}{\to} 0 \;\; a.s. \tag{3.5.45b}$$

Next, we only show (3.5.45a), from which relation (3.5.45b) follows immediately.

From (3.5.44b) and (3.5.44d) we have

$$\widehat{\theta}_t^{(t-1)}(L)\, \widehat{e}_t - \theta_t^*(L)\, e_t^* = \widehat{\epsilon}_t - \varepsilon_t,$$

and then

$$\widehat{\theta}_t^{(t-1)}\left(\widehat{e}_t - e_t^*\right) = \widehat{\epsilon}_t - \varepsilon_t + \left(\theta_t^*(L) - \widehat{\theta}_t^{(t-1)}\right) e_t^*.$$

Using again the fact that $\widehat{\beta}_t^{(t)}$ lies in the stability region, we have via Theorem 3.5.1 (see also (3.5.42))

$$\left|\widehat{e}_t - e_t^*\right| \le \sum_{j=1}^{t} \lambda^{t-j}\left(\left|\widehat{\epsilon}_t - \varepsilon_t\right| + j^{-1/2+\delta^*/2} o_j(1)\left|e_j^*\right|\right) + D\lambda^t,$$

whereupon the Toeplitz lemma ensures (3.5.45a), provided that we can show

$$\left|\widehat{\epsilon}_t - \varepsilon_t\right| \overset{t\to\infty}{\to} 0 \;\; \text{a.s.},$$

$$t^{-1/2+\delta^*/2}\left|e_t^*\right| \overset{t\to\infty}{\to} 0 \;\; \text{a.s.}$$

The first relation of these holds since it is a consequence of part $i)$ of Lemma 3.5.2, and the second from

$$\sup_{t\in\mathbb{Z}} E\left(\left\|\psi_t(\beta^*)\right\|^{4/1-3\delta^*}\right) < \infty, \tag{3.5.46}$$

in the same way that (3.5.43) follows from (3.5.34a). Note finally that (3.5.46) holds from assumption $B5$ and (3.5.35a) and hence from

$$\sup_{t\in\mathbb{Z}} E\left(\|\varphi_t\|^{4/1-3\delta^*}\right) < \infty,$$

in the same way that (3.5.34c) follows from (3.5.34a). This completes the proof. ■

Lemma 3.5.3

For a causal and periodically ergodic $PARMA$ process, suppose that the assumptions of Lemma 3.5.1 *hold, then*

$$\frac{1}{t}\widehat{R}_t \overset{t\to\infty}{\longrightarrow} \frac{1}{S}\sum_{s=1}^{S}\frac{1}{\sigma_s^2}E\left(\psi_s\left(\beta^*\right)\psi_s'\left(\beta^*\right)\right) \quad a.s. \tag{3.5.47}$$

Proof

Neglecting the initial values, we have from recursion (3.5.25d)

$$\begin{aligned}\frac{1}{t}\widehat{R}_t &= \frac{1}{t}\sum_{k=1}^{t}\frac{\widehat{\psi}_k\widehat{\psi}_k'}{\widehat{\sigma}_{i_k}^{2(i_k,\tau_k)}},\\ &= \frac{1}{t}\sum_{k=1}^{t}\frac{\widehat{\psi}_k\widehat{\psi}_k'}{\sigma_k^2}+\frac{1}{t}\sum_{k=1}^{t}\left(\frac{1}{\widehat{\sigma}_{i_k}^{2(i_k,\tau_k)}}-\frac{1}{\sigma_k^2}\right)\widehat{\psi}_k\widehat{\psi}_k',\end{aligned} \tag{3.5.48}$$

in the same way that we have established (3.3.15) (see Section 3.3) from $H1$ and $H2$, the following relation

$$\frac{1}{t}\sum_{k=1}^{t}\left(\frac{1}{\widehat{\sigma}_{i_k}^{2(i_k,\tau_k)}}-\frac{1}{\sigma_k^2}\right)\widehat{\psi}_k\widehat{\psi}_k' = o_p(1),$$

holds from (3.5.34c). Now the first term in the right member of (3.5.48) can be written as

$$\begin{aligned}\frac{1}{t}\sum_{k=1}^{t}\frac{\widehat{\psi}_k\widehat{\psi}_k'}{\sigma_k^2} &= \frac{1}{t}\sum_{k=1}^{t}\frac{\psi_k\left(\beta^*\right)\psi_k'\left(\beta^*\right)}{\sigma_k^2}+\frac{1}{t}\sum_{k=1}^{t}\frac{\left(\widehat{\psi}_k-\psi_k\left(\beta^*\right)\right)\widehat{\psi}_k'}{\sigma_k^2}\\ &\quad+\frac{1}{t}\sum_{k=1}^{t}\frac{\left(\widehat{\psi}_k-\psi_k\left(\beta^*\right)\right)\psi_k'\left(\beta^*\right)}{\sigma_k^2},\\ &= \frac{1}{t}\sum_{k=1}^{t}\frac{\psi_k\left(\beta^*\right)\psi_k'\left(\beta^*\right)}{\sigma_k^2}+A_t+B_t,\end{aligned}$$

in which the two last terms in the right member are weakly negligible, i.e.

$$A_t = \frac{1}{t}\sum_{k=1}^{t} \frac{\left(\widehat{\psi}_k - \psi_k\left(\beta^*\right)\right)\widehat{\psi}_k'}{\sigma_k^2} = o_p(1), \tag{3.5.49a}$$

$$B_t = \frac{1}{t}\sum_{k=1}^{t} \frac{\left(\widehat{\psi}_k - \psi_k\left(\beta^*\right)\right)\psi_k'\left(\beta^*\right)}{\sigma_k^2} = o_p(1). \tag{3.5.49b}$$

Indeed, from the Cauchy-Shwarz inequality, we have

$$\begin{aligned} \|A_t\| &\leq \frac{1}{t}\sum_{k=1}^{t} \frac{\left\|\widehat{\psi}_k - \psi_k\left(\beta^*\right)\right\| \left\|\widehat{\psi}_k'\right\|}{\sigma_k^2}, \\ &\leq \left[\left(\frac{1}{t}\sum_{k=1}^{t}\frac{1}{\sigma_k^2}\left\|\widehat{\psi}_k - \psi_k\left(\beta^*\right)\right\|^2\right)\left(\frac{1}{t}\sum_{k=1}^{t}\frac{1}{\sigma_k^2}\left\|\widehat{\psi}_k'\right\|^2\right)\right]^{1/2}. \end{aligned}$$

Since by Lemma 3.5.2, ii), we have $\left\|\widehat{\psi}_k - \psi_k\left(\beta^*\right)\right\| \overset{t\to\infty}{\longrightarrow} 0$, then by the Cesaro lemma (cf, Appendix B), we have

$$\frac{1}{t}\sum_{k=1}^{t}\frac{1}{\sigma_k^2}\left\|\widehat{\psi}_k - \psi_k\left(\beta^*\right)\right\|^2 \overset{t\to\infty}{\longrightarrow} 0,$$

and because of Lemma 3.5.1, iii), the periodic stationarity and then the periodic ergodicity of the $PARMA$ process, we have

$$\lim_{t\to\infty}\frac{1}{t}\sum_{k=1}^{t}\frac{1}{\sigma_k^2}\left\|\widehat{\psi}_k'\right\|^2 < \infty,$$

which implies (3.5.49a). Relation (3.5.49b) follows by the same argument.

Finally, to establish the lemma, it suffices to see that by periodic stationary and periodic ergodicity, we obtain for $t = s + S\tau$

$$\begin{aligned} \frac{1}{t}\sum_{k=1}^{t}\frac{\psi_k\left(\beta^*\right)\psi_k'\left(\beta^*\right)}{\sigma_k^2} &= \frac{1}{S}\sum_{s=1}^{S}\frac{1}{\tau}\sum_{m=0}^{\tau-1}\frac{\psi_{s+Sm}\left(\beta^*\right)\psi_{s+Sm}'\left(\beta^*\right)}{\sigma_k^2} + o_p(1), \\ &\overset{a.s}{\longrightarrow} \frac{1}{S}\sum_{s=1}^{S} E\left(\frac{\psi_s\left(\beta^*\right)\psi_s'\left(\beta^*\right)}{\sigma_k^2}\right). \end{aligned}$$

which completes the proof. ■

As mentioned above, to complete the proof, we have to show that the last term in (3.5.33) is negligible. Before doing this, we need to obtain the following result.

Lemma 3.5.4

Under assumption $B5$, we have

$$t^{1/2+\delta^*}\left(\widehat{\beta}_t - \widehat{\beta}_{t-1}\right) \overset{t\to\infty}{\to} 0,\ a.s., \tag{3.5.50}$$

where δ^* is the one given in $B5$.

Proof

From (3.5.25c) we have

$$\begin{aligned} t^{1/2+\delta^*}\left(\widehat{\beta}_t - \widehat{\beta}_{t-1}\right) &= t^{-1/2+\delta^*}\left(\frac{1}{t}\widehat{R}_t^{-1}\right)\widehat{\psi}_t\frac{\widehat{\varepsilon}_t}{\widehat{\sigma}_{i_t}^{2(i_t,\tau_t)}}, \\ &= t^{-1/2+\delta^*}\left(\frac{1}{t}\widehat{R}_t^{-1}\right)\widehat{\psi}_t\frac{\varepsilon_t}{\widehat{\sigma}_{i_t}^{2(i_t,\tau_t)}} + t^{-1/2+\delta^*}\left(\frac{1}{t}\widehat{R}_t^{-1}\right)\widehat{\psi}_t\frac{\widehat{\varepsilon}_t-\varepsilon_t}{\widehat{\sigma}_{i_t}^{2(i_t,\tau_t)}}. \end{aligned}$$

In view of Lemma 3.5.3 and the strong consistency of $\widehat{\sigma}_{i_t}^{2(i_t,\tau_t)}$, to establish the lemma, it is enough to show

$$t^{-1/2+\delta^*}\widehat{\psi}_t\varepsilon_t + t^{-1/2+\delta^*}\widehat{\psi}_t\left(\widehat{\varepsilon}_t - \varepsilon_t\right) \overset{t\to\infty}{\to} 0,\ a.s.,$$

which will follow if

$$t^{-1/4+\delta^*/2}\left|\varepsilon_t\right| \overset{t\to\infty}{\to} 0,\ a.s., \tag{3.5.51a}$$

$$t^{-1/4+\delta^*/2}\left|\widehat{\varepsilon}_t - \varepsilon_t\right| \overset{t\to\infty}{\to} 0,\ a.s., \tag{3.5.51b}$$

$$t^{-1/4+\delta^*/2}\left\|\widehat{\psi}_t\right\| \overset{t\to\infty}{\to} 0,\ a.s. \tag{3.5.51c}$$

By the Markov inequality and then the Borel-Cantelli lemma, assumption $B5$ ensures (3.5.51a). The same argument shows that (3.5.51c) follows from (3.5.34c) and that (3.5.51b) from

$$\sup_{t\in\mathbb{Z}} E\left(\left|\widehat{\varepsilon}_t - \varepsilon_t\right|^{4/1-3\delta^*}\right) < \infty,$$

which holds, via the C_r-inequality and Hölder inequality in the same way that (3.5.34b) holds. ■

Lemma 3.5.5

Under the assumptions of Lemma 3.5.1 *we have*

$$t^{-1/2}\sum_{k=1}^{t}\widehat{\psi}_k\frac{\left(\widehat{\psi}'_k\widetilde{\beta}_{k-1}+\widehat{\varepsilon}_k-\varepsilon_k\right)}{\widehat{\sigma}_{i_k}^{2(i_k,\tau_k)}}\overset{t\to\infty}{\to}0,\ a.s. \tag{3.5.52}$$

Proof

We first notice that the following proof of this lemma is very similar to the one given in Solo (1981, p. 315). Since $\widehat{\sigma}_{i_k}^{2(i_k,\tau_k)}$ is bounded a.s., (3.5.52) may be replaced by

$$t^{-1/2}\sum_{k=1}^{t}\left\|\widehat{\psi}_k\right\|\left|\widehat{\psi}'_k\widetilde{\beta}_{k-1}+\widehat{\varepsilon}_k-\varepsilon_k\right|\overset{t\to\infty}{\to}0,\ a.s.$$

Denote $h_k=\left|\widehat{\psi}'_k\widetilde{\beta}_{k-1}+\widehat{\varepsilon}_k-\varepsilon_k\right|$. Then it suffices to show that

$$\sum_{k=1}^{t}k^{-1/2}\left\|\widehat{\psi}_k\right\|\left|\widehat{\psi}'_k\widetilde{\beta}_{k-1}+\widehat{\varepsilon}_k-\varepsilon_k\right|<\infty, \tag{3.5.53}$$

and then (3.5.52) will follow from Kronecker's lemma (see Appendix B). Now since $\widehat{\psi}_k$ is obtained from $\widehat{\varphi}_t$ by an exponentially stable filter, it is easy to see that

$$\begin{aligned}\widehat{\theta}_t^{(t-1)}(L)\,\widehat{\psi}'_k\left(\widehat{\beta}_{k-1}-\beta^*\right) &= \widehat{\varphi}'_k\left(\widehat{\beta}_{k-1}-\beta^*\right)+O\left(\left\|\widehat{\beta}_k-\widehat{\beta}_{k-1}\right\|\left\|\widehat{\psi}_k\right\|\right), \\ &= \widehat{\varphi}'_k\left(\widehat{\beta}_{k-1}-\beta^*\right)+d_k,\end{aligned} \tag{3.5.54}$$

(see also Solo (1981, Appendix, p. 317). The derivation of (3.5.54) is tedious but simple. From (3.5.41a) and (3.5.54) we have

$$\widehat{\theta}_t^{(t-1)}(L)\,\widehat{\psi}'_k\left(\widehat{\beta}_{k-1}-\beta^*+(\widehat{\varepsilon}_k-\varepsilon_k)\right)=\left(\widehat{\theta}_t^{(t-1)}(L)-\theta_t^*(L)\right)(\widehat{\varepsilon}_k-\varepsilon_k)+d_k, \tag{3.5.55}$$

where in view of Lemma 3.5.4, (see (3.5.50)) we have

$$d_t=t^{-1/2-\delta^*}\left\|\widehat{\psi}_k\right\|.$$

Thus, from (3.5.55) and the assumption $B5$, we then have

$$\begin{aligned}
h_k &= \left|\widehat{\psi}_k' \widetilde{\beta}_{k-1} + \widehat{\varepsilon}_k - \varepsilon_k\right| \le \sum_{j=1}^{t} \lambda^{t-j} |d_j| + \sum_{j=1}^{t} \lambda^{t-j} \left\|\widehat{\varphi}_j\right\| |\widehat{\varepsilon}_k - \varepsilon_k| \, j^{-1/2-\delta^*} o_j(1), \\
&= \sum_{j=1}^{t} \lambda^{t-j} o_t(1) t^{-1/2-\delta^*} \left(\|\widehat{\varphi}_t\| + |\widehat{\varepsilon}_t - \varepsilon_t| \, t^{2\delta^*} \right), \\
&= \sum_{j=1}^{t} \lambda^{t-j} o_t(1) t^{-1/2-\delta^*} \gamma_t,
\end{aligned}$$

where $\gamma_t = \|\widehat{\varphi}_t\| + |\widehat{\varepsilon}_t - \varepsilon_t| \, t^{2\delta^*}$. Now, we have

$$\begin{aligned}
\sum_{k=1}^{t} k^{-1/2} \left\|\widehat{\psi}_k\right\| h_k &\le \sum_{k=1}^{t} k^{-1/2} \left\|\widehat{\psi}_k\right\| \sum_{j=1}^{k} \lambda^{k-j} o_k(1) k^{-1/2-\delta^*} \gamma_k, \\
&= \sum_{k=1}^{t} o_k(1) \gamma_k k^{-1/2-\delta^*} \sum_{j=k}^{t} \lambda^{k-j} \left\|\widehat{\psi}_j\right\| j^{1/2}, \\
&\le \sum_{k=1}^{t} k^{-1-\delta^*} \gamma_k o_k(1) \sum_{j=0}^{t-k} \lambda^{j} \left\|\widehat{\psi}_{j+k}\right\|, \\
&= \sum_{k=1}^{t} k^{-1-\delta^*} \gamma_k o_k(1) \varkappa_k,
\end{aligned}$$

where $\varkappa_k = \sum_{j=0}^{t-k} \lambda^{j} \left\|\widehat{\psi}_{j+k}\right\|$. Using (3.5.40$a$), to establish (3.5.53), we need only to show

$$\begin{aligned}
\sum_{k=1}^{\infty} \varkappa_k k^{-1-\delta^*} &< \infty, \\
\sum_{k=1}^{\infty} \varkappa_k \left\|\widehat{\psi}_k\right\| k^{-1-\delta^*} &< \infty,
\end{aligned}$$

which holds if

$$\begin{aligned}
\sum_{k=1}^{\infty} E\left(\varkappa_k\right) k^{-1-\delta^*} &< \infty, \\
\sum_{k=1}^{\infty} E\left(\varkappa_k \left\|\widehat{\psi}_k\right\|\right) k^{-1-\delta^*} &< \infty,
\end{aligned}$$

(see Lukacs, 1975, p. 80). These will hold if

$$\begin{aligned}\sup_t E\left(\varkappa_t\right) &< \infty, \\ \sup_t \left(\varkappa_t \left\|\widehat{\psi}_t\right\|\right) &< \infty.\end{aligned} \tag{3.5.56}$$

Now, part iii) of Lemma 3.5.1 ensure (3.5.56) via the Cauchy-Shwartz inequality. This completes the proof.

Proof of Theorem 3.5.2

Using the previous lemmas, equation (3.5.33) can be written as

$$t^{1/2}\left(\widehat{\beta}_t - \beta^*\right) = t^{-1/2}\left(\frac{1}{t}\widehat{R}_t\right)^{-1}\sum_{k=1}^{t}\widehat{\psi}_k \frac{\varepsilon_k}{\widehat{\sigma}_{i_k}^{2(i_k,\tau_k)}} + o_p(1).$$

To obtain result (3.5.32), we need to establish the following asymptotic convergence in law

$$t^{-1/2}\sum_{k=1}^{t}\widehat{\psi}_k \frac{\varepsilon_k}{\widehat{\sigma}_{i_k}^{2(i_k,\tau_k)}} \xrightarrow{D} N\left(0, \frac{1}{S}\sum_{s=1}^{S}\frac{1}{\sigma_k^2}E\left[\psi_s\left(\beta^*\right)\psi_s'\left(\beta^*\right)\right]\right).$$

But using similar arguments to those using to prove Lemma 3.5.3, we get

$$\begin{aligned}t^{-1/2}\sum_{k=1}^{t}\widehat{\psi}_k \frac{\varepsilon_k}{\widehat{\sigma}_{i_k}^{2(i_k,\tau_k)}} &= t^{-1/2}\sum_{k=1}^{t}\psi_k\left(\beta^*\right)\frac{\varepsilon_k}{\sigma_k^2} + t^{-1/2}\sum_{k=1}^{t}\left(\frac{1}{\widehat{\sigma}_{i_k}^{2(i_k,\tau_k)}} - \frac{1}{\sigma_k^2}\right)\psi_k\left(\beta^*\right)\varepsilon_k \\ &\quad + \sum_{k=1}^{t}\left(\widehat{\psi}_k - \psi_k\left(\beta^*\right)\right)\frac{\varepsilon_k}{\widehat{\sigma}_{i_k}^{2(i_k,\tau_k)}}, \\ &= t^{-1/2}\sum_{k=1}^{t}\psi_k\left(\beta^*\right)\frac{\varepsilon_k}{\sigma_k^2} + o_p(1).\end{aligned} \tag{3.5.57}$$

Thus, by the Martingale central limit theorem (cf, Hall and Heyde (1980, corollary 3.1)) which is easily applicable and the periodic ergodicity and periodic stationary we have

$$t^{-1/2}\sum_{k=1}^{t}\psi_k\left(\beta^*\right)\frac{\varepsilon_k}{\sigma_k^2} \xrightarrow{D} N\left(0, \frac{1}{S}\sum_{s=1}^{S}E\left(\frac{1}{\sigma_k^2}\left[\psi_s\left(\beta^*\right)\psi_s'\left(\beta^*\right)\right]\right)\right),$$

which achieves the proof.

3.6 Unified formulation of the $PELS$ and $PRML$ algorithms : A summary

In the above, we have seen that the $PELS$ and the $PRML$ are obtained from different approaches, even though they have a similar form. Indeed, the $PELS$ belongs to the class of pseudo-linear regression methods, while the $PRML$ can be seen as a recursive predictor error method. The distinction between these two algorithms is motivated by their different numerical and statistical properties. The purpose of this section is to show that both algorithms can be formulated via a single set of recursions through a certain prefilter. Moreover, it will be shown that with a modified version of this prefilter, the resulting algorithm lies somewhere in between the $PELS$ an the $PRML$. Similar ideas were already considered by Friedlander (1982) and Goodwin and Sin (1984) for $ARMAX$ recursive estimation.

Consider the following $PARMA$ model written in a pseudo-regression form

$$y_t = \varphi_t' \beta + \varepsilon_t, \tag{3.6.1}$$

where $\{\varepsilon_t,\ t \in \mathbb{Z}\}$ is an S-periodic white noise with variance σ_t^2 and where

$$\varphi_{i+S\tau} = \Big(\underbrace{\mathbf{0}', ..., \mathbf{0}'}_{i-1 \text{ time}}, y_{i+S\tau-1}, ..., y_{i+S\tau-p}, -\varepsilon_{i+S\tau-1}, ..., -\varepsilon_{i+S\tau-q}, \underbrace{\mathbf{0}', ..., \mathbf{0}'}_{S-i \text{ time}} \Big)', \tag{3.6.2}$$

with $\mathbf{0}$ being the $(p+q) \times 1$ null vector, $i = 1, 2, ..., S$ and $\tau \in \mathbb{Z}$. Now, a general algorithm is proposed

Algorithm 3.6.1

A general recursive parameter estimation algorithm for $PARMA$ models is given by the follow-

ing recursion

$$\widehat{\varepsilon}_{i+S\tau} = y_{i+S\tau} - \widehat{\beta}'_{i+S\tau-1}\widehat{\varphi}_{i+S\tau} \text{ estimated prediction error (a priori),} \tag{3.6.3a}$$

$$\widehat{\sigma}_i^{2(i,\tau)} = \begin{cases} \dfrac{\tau}{\tau+1}\widehat{\sigma}_i^{2(i,\tau-1)} + \dfrac{1}{\tau+1}\widehat{\varepsilon}^2_{i+S\tau}, \ 0<\tau\le\delta, \\ \dfrac{\tau-\delta}{\tau+1-\delta}\widehat{\sigma}_i^{2(i,\tau-1)} + \dfrac{1}{\tau+1-\delta}\widehat{\varepsilon}^2_{i+S\tau}, \ \tau>\delta, \end{cases} \tag{3.6.3b}$$

$$\widehat{\beta}_{i+S\tau} = \widehat{\beta}_{i+S\tau-1} + \frac{\widehat{P}_{i+S\tau}\widehat{\psi}_{i+S\tau}\widehat{\varepsilon}_{i+S\tau}}{\widehat{\sigma}_i^{2(i,\tau)} + \widehat{\psi}'_{i+S\tau}\widehat{P}_{i+S\tau-1}\widehat{\psi}_{i+S\tau}}, \ \widehat{\beta}_0 = 0, \tag{3.6.3c}$$

$$\widehat{P}_{i+S\tau} = \widehat{P}_{i+S\tau-1} - \frac{\widehat{P}_{i+S\tau-1}\widehat{\psi}_{i+S\tau}\widehat{\psi}'_{i+S\tau}\widehat{P}_{i+S\tau-1}}{\widehat{\sigma}_i^{2(i,\tau)} + \widehat{\psi}'_{i+S\tau}\widehat{P}_{i+S\tau-1}\widehat{\psi}_{i+S\tau}}, \ \widehat{P}_0 = K.I, \tag{3.6.3d}$$

$$\widehat{\psi}_{t+1} = \sum_{j=1}^{q}\left(M^j_{i,1}\widehat{\theta}^{(t)}_{t+1,j} + M^j_{i,2}\widetilde{\theta}_{t,j}\right)\widehat{\psi}_{t+1-j} + \widehat{\varphi}_{t+1}, \ (t = i+S\tau), \tag{3.6.3e}$$

$$\widehat{\epsilon}_{i+S\tau} = y_{i+S\tau} - \widehat{\beta}'_{i+S\tau}\widehat{\varphi}_{i+S\tau}, \qquad \text{estimated residual (a posteriori),} \tag{3.6.3f}$$

$$\text{Compute } \widehat{\varphi}_{i+S\tau} \text{ from (3.6.2) with } \widehat{\epsilon}_{i+S\tau} \text{ in place of } \varepsilon_{i+S\tau}, \tag{3.6.3g}$$

where $\delta = (p+q)\,S$, K *is a large enough number,* $0 \le M_{i,j} \le 1$,$j = 1,2$, $i = 1,...,S$ *and* $\widehat{\sigma}_i^{2(i,0)} = \widehat{\varepsilon}_i^2$.

We note that the key step in the latter algorithm is the prefilter equation (3.6.3e), i.e.

$$D_t\,(L)\,\widehat{\psi}_t = \widehat{\varphi}_t, \tag{3.6.4}$$

where

$$D_t\,(L) = \widehat{\theta}^{(t)}_t\,(M_{i_t,1}L) + \widetilde{\theta}_t\,(M_{i_t,2}L)\,,$$

with $\widehat{\theta}^{(t)}_t\,(M_{i_t,1}L) = 1 - \sum_{j=1}^{q}\widehat{\theta}^{(t)}_{t,j}L^j$ and so on. Any particular value of the "parameter prefilter" given by (3.6.4) specifies a particular algorithm among the last ones studied in the above. We have indeed,

i) for $M_{i,1} = 0$ and $M_{i,2} = 0$ for all $i \in \{1,...,S\}$ Algorithm 3.6.1 reduces to the $PELS$ algorithm without prefilter (Algorithm 3.4.1);

ii) when $M_{i,1} = 0$ and $M_{i,2} = 1$ for all $i \in \{1,...,S\}$, Algorithm 3.6.1 reduces to the $PELS$ algorithm with prefilter (Algorithm 3.4.2);

iii) let $M_{i,1} = 1$ and $M_{i,2} = 0$ for all $i \in \{1, ..., S\}$, gives the $PRML$ algorithm (Algorithm 3.5.1, given by (3.5.26)).

The convergence of the estimates provided by this algorithm is related to the positive realness of the transfer functions $H_t(L)$ where

$$H_t(L) = \frac{D_t(L)}{\theta_t(L)} - \frac{1}{2}, \quad t \in \{1, ..., S\}. \tag{3.6.5}$$

The $PELS$ algorithm (Algorithm 3.4.1) will not always converge since $H_t(L)$ is not positive real for a certain choice of $\theta_t(L)$, even if this last is stable, i.e., the $PARMA$ model is invertible. This is due to the fact that the positive realness condition is more stringent than the stability condition as we have seen in the above (see Remark 3.4.5). The inclusion of a prefilter in the $PELS$ (see Algorithm 3.4.2) can then relax the restrictive positive realness condition, while increasing the algorithmic complexity. A common limitation of the $PELS$ type algorithms (with or without prefilter) is the asymptotic inefficiency of the parameter estimates. The $PRML$ algorithm, on the other hand is shown to produce asymptotically efficient estimates. Moreover, the choice of $D_t(L) = \widehat{\theta}_t^{(t)}(L)$ can guarantee the positive realness of $H_t(L)$ provided that $\theta_t(L)$ is stable. However, unlike the $PELS$ algorithms, this algorithm requires monitoring the stability of the $\widehat{\theta}_t^{(t)}(L)$ polynomials and then the projection of the parameter estimate into a stable region. This can be very costly, and then there is interest too in less efficient recursions without monitoring procedures.

Finally we note that the inclusion of the constants M_{i1} and M_{i2} (such that $0 \leq M_{i,j} \leq 1$) in the prefilter $D_t(L)$ is motivated, in the case when the zeros of the $\theta_t(L)$ are close to the unit circle, by the fact that if z^* is a root of $D\left(z^{-1}\right)$ then Mz^* is a root of $D\left(Mz^{-1}\right)$, and hence for $M_{ij} < 1$ the roots Mz^{-1} will be strictly withing the stability region. Thus for $0 < M_{ij} < 1$ the obtained algorithm has properties that lie in between those of the $PELS$ and the $PRML$.

3.7 Periodic recursive Fisher-Scoring algorithm for PARMA models

3.7.1 The P-RFS algorithm

From a theoretical viewpoint, under certain assumptions the preceding algorithms give consistent estimators. In particular the obtained $PRML$ estimator is asymptotically efficient and then it has the same asymptotic properties as its off-line homologue, the ML estimator (and then the WLS estimator, see Section 1.6). However, for a finite sample size, and as is well known in the time-invariant recursive estimation literature (e.g. Ljung and Söderström, 1983, p. 195) the off-line ML estimate might be strictly better than the on-line one (i.e. our $PRML$). This is mainly because the $PRML$ algorithm optimizes an approximate ML criterion instead of the exact one. Furthermore, the approximate Hessian denoted by $\widehat{R}_t$ is extremely variable and might cause undesirable disturbances for the current estimate, especially for the transient stage, before that the estimate reaches to a "small" neighborhood of the true value. Since according to Lemma 3.5.3 the approximate covariance matrix $\frac{1}{t}\widehat{R}_t$ converges to the asymptotic Fisher information matrix $F\left(\beta^*, \boldsymbol{\sigma}^*\right)$ (multiplied by $\frac{1}{S}$ and evaluated at the true value of the parameter), the replacement of $\frac{1}{t}\widehat{R}_t$ by its limit $F\left(\beta^*, \boldsymbol{\sigma}^*\right)$ which is also the mean of the likelihood Hessian, will stabilize the estimate. However, as the true values β^* and $\boldsymbol{\sigma}^*$ are unknown, the expression $F\left(\widehat{\beta}_t, \widehat{\boldsymbol{\sigma}}_t\right)$ will be a good approximation of $F\left(\beta^*, \boldsymbol{\sigma}^*\right)$ when $\widehat{\beta}_t$ and $\widehat{\boldsymbol{\sigma}}_t$ are consistent estimates of β^* and $\boldsymbol{\sigma}^*$, respectively. In this section we propose to slightly modify the $PRML$ algorithm in order to improve its finite properties. We replace in the updating equation (3.5.25c) of the $PRML$ algorithm, the matrix $\widehat{R}_t$ which is an approximate of the Hessian $\left(\frac{\partial^2 V_t(\beta)}{\partial\beta\partial\beta'}\right)$ by its expectation $F_t = \frac{t-1}{S} F\left(\widehat{\beta}_{t-1}, \widehat{\boldsymbol{\sigma}}_{t-1}\right)$ evaluated at $\widehat{\beta}_{t-1}$ and $\widehat{\boldsymbol{\sigma}}_t$, where $\widehat{\boldsymbol{\sigma}}_t = \left(\widehat{\sigma}_1^{2(t)}, ..., \widehat{\sigma}_S^{2(t)}\right)$. This leads to a recursive analogue of the well-known Fisher-Scoring algorithm (hence the name), devoted to the off-line minimization (see e.g. Sen and Singer 1993, p. 205).

For the on-line case, the same idea was considered by Duflo (1990) for ($i.i.d.$) sequence and by Zahaf (1998) for $ARMA$ recursive estimation (the obtained algorithm was called RML_{MZ}, see also Klein and Mélard, 2004). Thus the updating equation (3.5.25c) for the recursive estimate $\widehat{\beta}_t$ becomes

$$\widehat{\beta}_t = \widehat{\beta}_{t-1} + F_t^{-1}\widehat{\psi}_t\widehat{\varepsilon}_t, \quad \widehat{\beta}_0 = 0, \tag{3.7.1}$$

where $F_t = \frac{t-1}{S} F\left(\widehat{\beta}_{t-1}, \widehat{\boldsymbol{\sigma}}_t\right)$ is calculated from (1.4.8) (see Section 1.4). The recursive calculation of $\widehat{R}_t$ in (3.5.25) is however useless and it will thus be removed. Using the last considerations, we obtain the following recursive estimation algorithm for $PARMA$ models.

Algorithm 3.7.1

The recursive Fisher-Scoring algorithm for $PARMA$ recursive estimation is given by the following recursions

$$\widehat{\varepsilon}_{i+S\tau} = y_{i+S\tau} - \widehat{\beta}'_{i+S\tau-1}\widehat{\varphi}_{i+S\tau} \text{ estimated prediction error (a priori),} \quad (3.7.2a)$$

$$\widehat{\sigma}_i^{2(i,\tau)} = \begin{cases} \dfrac{\tau}{\tau+1}\,\widehat{\sigma}_i^{2(i,\tau-1)} + \dfrac{1}{\tau+1}\widehat{\varepsilon}^2_{i+S\tau}, & 0 < \tau \leq \delta, \\ \dfrac{\tau-\delta}{\tau+1-\delta}\,\widehat{\sigma}_i^{2(i,\tau-1)} + \dfrac{1}{\tau+1-\delta}\widehat{\varepsilon}^2_{i+S\tau}, & \tau > \delta, \end{cases} \quad (3.7.2b)$$

$$F_{i+S\tau} = \frac{i+S\tau-1}{S}F\left(\widehat{\beta}_{i+S\tau-1}, \widehat{\boldsymbol{\sigma}}_{i+S\tau}\right), \text{ (from (1.4.8)),} \quad (3.7.2c)$$

$$\widehat{\beta}_{i+S\tau} = \widehat{\beta}_{i+S\tau-1} + F^{-1}_{i+S\tau}\widehat{\psi}_{i+S\tau}\widehat{\varepsilon}_{i+S\tau}, \qquad \widehat{\beta}_0 = 0, \quad (3.7.2d)$$

$$\widehat{\psi}_{t+1} = \sum_{j=1}^{q}\widehat{\theta}^{(t)}_{t+1,j}\widehat{\psi}_{t+1-j} + \widehat{\varphi}_{t+1}, \; (t = i+S\tau), \quad (3.7.2e)$$

$$\widehat{\epsilon}_{i+S\tau} = y_{i+S\tau} - \widehat{\beta}'_{i+S\tau}\widehat{\varphi}_{i+S\tau} \qquad \text{estimated residual (a posteriori),} \quad (3.7.2f)$$

$$\text{Compute } \widehat{\varphi}_{i+S\tau} \text{ from (3.6.2) with } \widehat{\epsilon}_{i+S\tau} \text{ in place of } \varepsilon_{i+S\tau}, \quad (3.7.2g)$$

with the starting up values $\widehat{\sigma}_i^{2(i,0)} = \widehat{\varepsilon}_i^2$, $i = 1, ...S$.

Remark 3.7.1

While the P-RFS algorithm seems to improve the accuracy of the $PRML$ estimate (see the following section) it has however less computational efficiency than the $PRML$ since on the one hand, it requires the inversion of the Fisher information matrix F_t, which has no Toeplitz structure in our periodic case, and on the other hand no good devices such as the matrix inversion lemma or the square root technique (cf, Breiman, 1977) may be used. Indeed, for the standard $ARMA$ case, Zahaf (1998) has exploited the Toeplitz structure of the Fisher information matrix for $ARMA$ models to develop a computationally simple method for inverting the Fisher information matrix (e.g. Rissanen 1973). This method was shown to be comparable to the square root algorithm (cf, Breimann, 1977) for updating and inverting $\widehat{R}_t$. In our periodic case, the Fisher information matrix F_t does not have a Toeplitz structure and we cannot alleviate the computational cost involved in the calculation of F_t^{-1}.

Remark 3.7.2

Notice that since the $P\text{-}RFS$ algorithm is a particular variant of the $PRML$, it must include a monitoring procedure to prevent the estimate from being outside the admissibility region. Thus we can apply the same monitoring procedure as that developed for the $PRML$ (see the Monitoring algorithm, Section 3.5), except that, in addition to the stability test of the estimated PAR and PMA polynomials, we must also test the redundancy of the estimated $PARMA$ model, in order to keep the Fisher information matrix invertible (see Theorem 1.4.1).

Remark 3.7.3

Zahaf (1998) has studied the statistical properties of the RML_{MZ} algorithm based on the work of Duflo (1990). The strong consistency and asymptotic normality of the RML_{MZ} estimates were established on the basis of the Robbins and Monro (1951) theorem.

3.8 Simulation study

For the implementation of the above recursive algorithms, we mention the need to check, at each iteration, the causality and invertibility of the obtained estimates and the algorithm robustness with respect to the starting parameter values. To take into account these two points, we have applied the procedures proposed by Ljung and Söderström (1983, p. 366), which tests the causality of the current estimate and, in the case of rejection, project it into the causal domain. As mentioned above, such a projection is implemented as successive reductions of the correction term for the parameter estimates. Furthermore, the causality condition is checked at each iteration by using the techniques given in Section 1.2. In this section we study the performance of some of the proposed algorithms via a simulation study. In fact, the lack of space prevents us being studying the finite properties of all the proposed algorithms.

3.8.1 Finite properties of the $PRLS$ algorithm

The performance of the proposed $PRLS$ algorithm was assessed by means of an intensive simulation study. Indeed, our algorithm is applied to periodic autoregressive processes of different orders ($PAR(1)$, $PAR(2)$ and $PAR(3)$), periods (4 and 7) and for a variety of sample sizes (T). For each model, we have considered one thousand Monte Carlo replications. The autoregressive parameters and the variances for each season are given, respectively, by Table $A1.4$, Table $A1.7$, Table $A2.4$, Table $A2.7$, Table $A3.4$ and Table $A3.7$. For all simulations, the unobservable starting values

$\underline{y}'_{1,0} = (y_0, y_{-1}, ..., y_{1-p})'$ of the stochastic difference equation (3.3.1) and the initial parameters estimates are set as zero and $P_0 = 100000I$.

Data generator processes

The considered models and their parameters values are :

Model 1. $PAR_S(1)$: $y_{s+S\tau} = \phi_s\, y_{s+S\tau-1} + \varepsilon_{s+S\tau}$, where $\varepsilon_{s+S\tau} \sim N(0,\, \sigma_s^2)$, $s = 1, ..., S$, $\tau \in \mathbb{Z}$;

$\underline{\phi} \diagdown s$	1	2	3	4
ϕ_s	$-.5$	.8	2	.35
σ_s^2	1	2	3	4

Table $A1.4$

1	2	3	4	5	6	7
1.5	.8	2	.7	1.2	.5	1
1	2	3	4	5	6	7

Table $A1.7$

Model 2. $PAR_S(2)$: $y_{s+S\tau} = \phi_{s,1}\, y_{s+S\tau-1} + \phi_{s,2}\, y_{s+S\tau-2} + \varepsilon_{s+S\tau}$, where $\varepsilon_{s+S\tau} \sim N(0,\, \sigma_s^2)$, $s = 1, ..., S$, $\tau \in \mathbb{Z}$;

$\underline{\phi} \diagdown s$	1	2	3	4
$\phi_{s,1}$	$-.5$	1.9	$-.2$	-2
$\phi_{s,2}$	$-.75$	.84	$-.48$	.48
σ_s^2	1	2	3	0.5

Table $A2.4$

1	2	3	4	5	6	7
$-.2$	1.2	$-.3$	1.7	1	.24	-1.4
.35	$-.27$	.54	$-.7$	$-.24$	.20	$-.45$
1	2	.5	3	4	10	.1

Table $A2.7$

Model 3.$PAR_S(3)$: $y_{s+S\tau} = \phi_{s,1} y_{s+S\tau-1} + \phi_{s,2}\, y_{s+S\tau-2} + \phi_{s,3}\, y_{s+S\tau-3} + \varepsilon_{s+S\tau}$, where $\varepsilon_{s+S\tau} \sim N(0, \sigma_s^2)$, $s = 1, ..., S, \tau \in \mathbb{Z}$;

$\underline{\phi} \diagdown s$	1	2	3	4
$\phi_{s,1}$	.43	.5	$-.3$	.45
$\phi_{s,2}$	1.4	$-.4$	.6	$-.1$
$\phi_{s,3}$	.6	1.2	$-.8$	.5
σ_s^2	1	2	3	0.5

Table $A3.4$

1	2	3	4	5	6	7
.43	.5	$-.3$	.45	$-.3$	.25	$-.7$
1.4	$-.4$	.6	$-.1$	.7	$-.34$	.65
.6	1.2	$-.8$	.5	-1.5	.8	1.1
1	1	1	1	1	1	1

Table $A3.7$

Simulation results and comments

Our recursive algorithm is applied to obtain parameter estimates for several models given by Tables $Ap.d$, for $p = 1, 2, 3$ and $d = 4, 7$. We have selected the parameters in order to remain compatible with Boshnakov (1996). The sample size was varied between 100 and 400 according to the orders of the considered models, to obtain a good appreciation of the aforementioned algorithm when applied to time series with a short and moderate size. Table $Bp.d$ shows means and, in parentheses, standard deviations of the obtained estimates corresponding to the $PAR_d\,(p)$ model, for the 1000 replications. In addition, Figure $1.s$, $s = 1, ..., 4$, and Figure $2.s$, $s = 1, ..., 4$, illustrate the behavior of the autoregressive parameter means and their standard deviations and the innovation variance means and their standard deviations, respectively, for the $PAR_4(1)$ model given by Table $A1.4$. One sees that the parameters are fairly well estimated even for small sample sizes and the obtained $PRLS$ estimates perform better than those obtained in Boshnakov's (1996) algorithm (noted B*, in Table $B3.7$). Moreover, our algorithm is of an on-line nature unlike that of Boshnakov (1996).

$\widehat{\underline{\phi}} \diagdown s$	1	2	3	4
$\widehat{\phi}_{s,1}^{(100)}$	$-.4992$ (.0843)	.7836 (.1861)	2.0036 (.1992)	.3463 (.0955)
$\widehat{\sigma}_s^{2(100)}$	1.0532 (.2167)	2.0788 (.4484)	3.1416 (.7291)	4.1838 (.8996)
$\widehat{\phi}_{s,1}^{(200)}$	$-.5026$ (.0577)	.7918 (.1258)	1.9995 (.1336)	.3467 (.0670)
$\widehat{\sigma}_s^{2(200)}$	1.0363 (.1813)	2.0495 (.3639)	3.1178 (.5550)	4.1294 (.6920)
$\widehat{\phi}_{s,1}^{(400)}$	$-.5039$ (.0412)	.7922 (.0894)	2.0001 (.0968)	.3476 (.0456)
$\widehat{\sigma}_s^{2(400)}$	1.0283 (.1518)	2.0462 (.3124)	3.0922 (.4670)	4.0923 (.5968)

Table $B1.4$, $PAR_4(1)$

$\underline{\widehat{\phi}} \diagdown s$	1	2	3	4
$\widehat{\phi}_{s,1}^{(100)}$	$-.4768$ (.2161)	1.8247 (.5175)	$-.1991$ (.1019)	-1.8750 (.3199)
$\widehat{\phi}_{s,2}^{(100)}$	$-.7181$ (.3381)	.8381 (.2050)	$-.4624$ (.2641)	.4615 (.0916)
$\widehat{\sigma}_s^{2(100)}$	1.0565 (.2321)	2.2241 (.5363)	3.2020 (.7201)	.5887 (.1917)
$\widehat{\phi}_{s,1}^{(200)}$	$-.4857$ (.1462)	1.8652 (.2565)	$-.2018$ (.0856)	-1.9284 (.2122)
$\widehat{\phi}_{s,2}^{(200)}$	$-.7293$ (.2409)	.8405 (.1070)	$-.4600$ (.1780)	.4690 (.0546)
$\widehat{\sigma}_s^{2(200)}$	1.0462 (.1763)	2.1798 (.3836)	3.1785 (.5135)	.5656 (.1413)
$\widehat{\phi}_{s,1}^{(400)}$	$-.4943$ (.1074)	1.8853 (.1572)	$-.2004$ (.0595)	-1.9651 (.1187)
$\widehat{\phi}_{s,2}^{(400)}$	$-.7439$ (.1825)	.8398 (.0590)	$-.4740$ (.1206)	.4736 (.0324)
$\widehat{\sigma}_s^{2(400)}$	1.0271 (.1296)	2.1507 (.2785)	3.1056 (.4187)	.5460 (.0936)

Table $B2.4$, $PAR_4(2)$

$\underline{\widehat{\phi}} \diagdown s$	1	2	3	4
$\widehat{\phi}_{s,1}^{(200)}$	.4417 (.1328)	.5162 (.1178)	$-.2988$ (.1165)	.4415 (.0537)
$\widehat{\phi}_{s,2}^{(200)}$	1.3161 (.1195)	$-.4647$ (.1428)	.5591 (.1189)	$-.1096$ (.0364)
$\widehat{\phi}_{s,3}^{(200)}$	.5824 (.0449)	1.1250 (.1952)	$-.6684$ (.1773)	.4948 (.0444)
$\widehat{\sigma}_s^{2(200)}$	1.1100 (.1870)	2.2839 (.4206)	3.3836 (.5999)	.5411 (.1001)
$\widehat{\phi}_{s,1}^{(300)}$	.4290 (.0830)	.5140 (.1071)	$-.3099$ (.0965)	.4507 (.0434)
$\widehat{\phi}_{s,2}^{(300)}$	1.2971 (.0685)	$-.4512$ (.1116)	.5513 (.1091)	$-.0988$ (.0372)
$\widehat{\phi}_{s,3}^{(300)}$	.5805 (.0370)	1.1439 (.1624)	$-.6901$ (.1751)	.4960 (.0327)
$\widehat{\sigma}_s^{2(300)}$	1.0989 (.1645)	2.2467 (.3520)	3.2995 (.5152)	.5363 (.0826)
$\widehat{\phi}_{s,1}^{(400)}$	.4343 (.0696)	.5294 (.1034)	$-.2690$ (.0906)	.4502 (.0376)
$\widehat{\phi}_{s,2}^{(400)}$	1.3530 (.0578)	$-.4349$ (.1064)	.5354 (.1144)	$-.0935$ (.0283)
$\widehat{\phi}_{s,3}^{(400)}$	.5908 (.0384)	1.1252 (.1445)	$-.7280$ (.1363)	.5024 (.0206)
$\widehat{\sigma}_s^{2(400)}$	1.0944 (.1468)	2.1279 (.2995)	3.1926 (.4325)	.5159 (.0756)

Table $B3.4$, $PAR_4(3)$

$\widehat{\underline{\phi}} \diagdown s$	1	2	3	4	5	6	7
$\widehat{\phi}_{s,1}^{(140)}$	1.4971 (.0174)	.7955 (.0142)	1.9990 (.0103)	.6972 (.0111)	1.1941 (.0215)	.4861 (.0277)	.9814 (.0251)
$\widehat{\sigma}_s^{2(140)}$	0.9128 (.1792)	2.1239 (.4506)	3.4451 (.6008)	3.7180 (.7521)	4.7392 (1.0455)	5.9821 (.6881)	7.5237 (1.8972)
$\widehat{\phi}_{s,1}^{(280)}$	1.4976 (.0088)	.7967 (.0084)	1.9992 (.0066)	.6980 (.0064)	1.1970 (.0094)	.4903 (.0169)	.9912 (.0129)
$\widehat{\sigma}_s^{2(280)}$	1.0013 (.1643)	2.0685 (.3900)	3.3463 (.5698)	3.8651 (.5018)	4.8111 (.9293)	6.0907 (.6007)	7.1031 (1.4076)
$\widehat{\phi}_{s,1}^{(420)}$	1.4982 (.0059)	.7972 (.0058)	1.9992 (.0039)	.6982 (.0044)	1.1976 (.0069)	.4926 (.0113)	.9949 (.0078)
$\widehat{\sigma}_s^{2(420)}$	1.0013 (.1378)	2.0145 (.3882)	3.1455 (.4613)	4.0979 (.4736)	5.0473 (.6349)	6.0864 (.5675)	7.0881 (1.0959)

Table $B1.7$, $PAR_7(1)$

$\widehat{\underline{\phi}} \diagdown s$	1	2	3	4	5	6	7
$\widehat{\phi}_{s,1}^{(140)}$	−.1975 (.1267)	1.2276 (.3188)	−.3076 (.1085)	1.6757 (.4757)	.9917 (.6587)	.2176 (.1746)	−1.3988 (.0231)
$\widehat{\phi}_{s,2}^{(140)}$	.3430 (.2168)	−.2515 (.1364)	.5534 (.2040)	−.6941 (.0815)	−.2603 (.1263)	.1834 (.1086)	−.4480 (.0190)
$\widehat{\sigma}_s^{2(140)}$	1.0937 (.3204)	2.3589 (.6317)	.5692 (.2476)	3.1595 (1.0425)	4.3596 (1.0955)	10.4457 (3.3437)	.1129 (.0328)
$\widehat{\phi}_{s,1}^{(210)}$	−.1977 (.0995)	1.2195 (.2545)	−.3112 (.0907)	1.6745 (.4274)	.9942 (.0998)	.2206 (.1377)	−1.4001 (.0182)
$\widehat{\phi}_{s,2}^{(210)}$	.3520 (.1710)	−.2558 (.1064)	.5615 (.1727)	−.6996 (.0627)	−.2928 (.1032)	.1877 (.0846)	−.4483 (.0160)
$\widehat{\sigma}_s^{2(210)}$	1.0716 (.2535)	2.2499 (.5642)	.5412 (.1274)	3.0895 (.9491)	4.2046 (.9715)	10.1347 (2.3780)	.1063 (.0246)
$\widehat{\phi}_{s,1}^{(280)}$	−.2016 (.0859)	1.2085 (.2126)	−.3081 (.0812)	1.6750 (.3680)	.9963 (.0862)	.2202 (.0969)	−1.4005 (.0157)
$\widehat{\phi}_{s,2}^{(280)}$	.3519 (.1497)	−.2593 (.0894)	.5550 (.1488)	−.7005 (.0553)	−.2559 (.0570)	.1928 (.0761)	−.4486 (.0133)
$\widehat{\sigma}_s^{2(280)}$	1.0605 (.2139)	2.2073 (.4654)	.5343 (.1077)	3.0789 (.7405)	4.1850 (.8404)	10.0696 (2.0271)	.1067 (.0209)

Table $B2.7$, $PAR_7(2)$

$\widehat{\underline{\phi}} \diagdown s$	1	2	3	4	5	6	7
$\widehat{\phi}^{(140)}_{s,1}$	.4336 (.0813)	.5950 (.1650)	−.2950 (.1138)	.4458 (.1118)	−.3015 (.1128)	.2493 (.0484)	−.6960 (.1256)
$\widehat{\phi}^{(140)}_{s,2}$	1.4091 (.1239)	−.4393 (.2311)	.6021 (.0904)	−.0933 (.0875)	.6965 (.1330)	−.3241 (.1246)	.6472 (.0569)
$\widehat{\phi}^{(140)}_{s,3}$	.5964 (.0290)	1.2938 (.2505)	−.7959 (.1061)	.4949 (.0669)	−1.4941 (.0740)	.7895 (.1232)	1.0977 (.0949)
$\widehat{\sigma}^{2(140)}_{s}$	1.1299 (.2330)	1.0280 (.0721)	1.0954 (.2607)	1.1786 (.2513)	1.1259 (.2751)	1.1240 (.3999)	1.0594 (.1983)
$\widehat{\phi}^{(210)}_{s,1}$	.4309 (.0562)	.5532 (.1420)	−.2984 (.0811)	.4486 (.0856)	−.2995 (.0851)	.2498 (.0353)	−.7038 (.0945)
$\widehat{\phi}^{(210)}_{s,2}$	1.4067 (.0917)	−.3883 (.0975)	.5998 (.0656)	−.0968 (.0672)	.6976 (.0930)	−.3363 (.0902)	.6501 (.0433)
$\widehat{\phi}^{(210)}_{s,3}$	.5982 (.0215)	1.2593 (.1440)	−.7966 (.0752)	.4978 (.0509)	−1.4977 (.0559)	.7937 (.0928)	1.0996 (.0691)
$\widehat{\sigma}^{2(210)}_{s}$	1.0391 (.1618)	1.0163 (.1288)	1.0533 (.1686)	1.0488 (.1828)	1.1050 (.2741)	1.0466 (.2238)	1.0395 (.1753)
$\widehat{\phi}^{(B^*)}_{s,1}$	.4471 (.0984)	.5388 (.0502)	−.2520 (.0914)	.4577 (.0837)	−.2570 (.3040)	.2503 (.0369)	−.6547 (.1182)
$\widehat{\phi}^{(B)}_{s,2}$	1.4656 (.1805)	−.4116 (.0718)	.5651 (.0732)	−.0749 (.0734)	.6550 (.1882)	0.3147 (.0943)	.5947 (.0689)
$\widehat{\phi}^{(B)}_{s,3}$	.5555 (.0526)	1.1775 (.1877)	−.7542 (.0838)	.4755 (.0561)	−1.3727 (.1632)	.7661 (.0980)	1.0732 (.1077)
$\widehat{\sigma}^{2(B)}_{s}$	1.1660 (.2081)	1.1337 (.1908)	1.0727 (.1637)	.9913 (.2620)	1.2111 (.2499)	.9871 (.2822)	1.1136 (.1812)

Table $B3.7$, $PAR_7(3)$, B* : Boshnakov (1996)

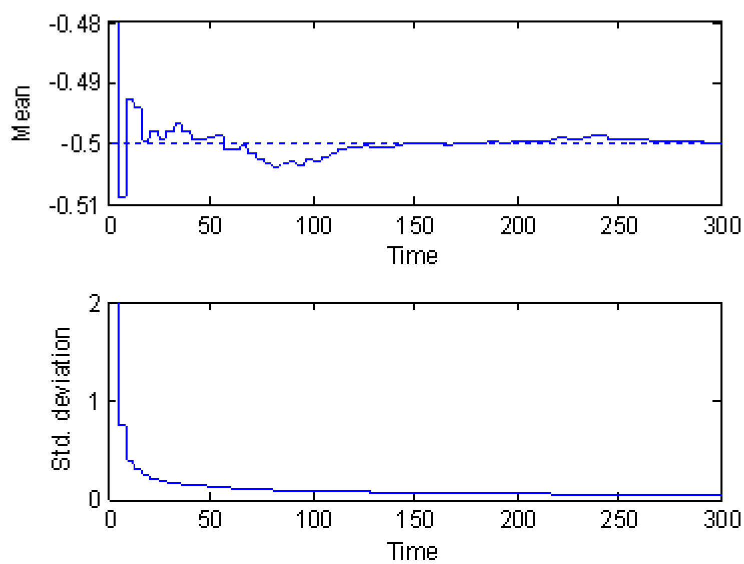


Figure 1.1 : $\widehat{\phi}^{(t)}_{1,1}$

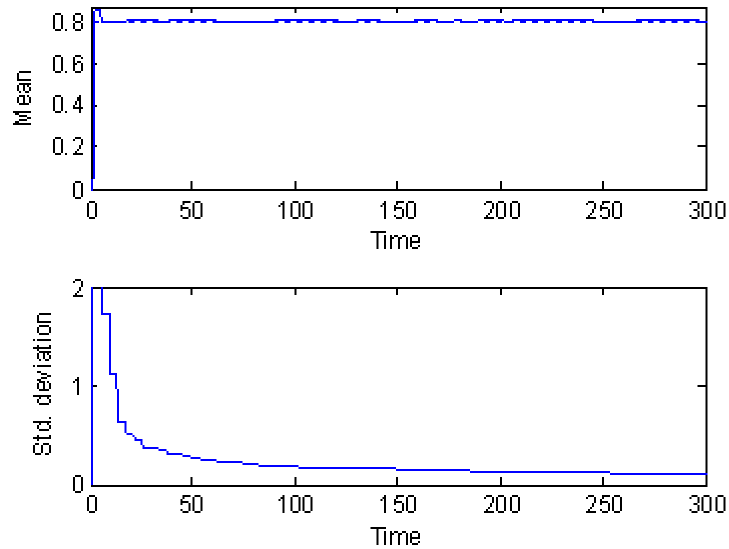


Figure 1.2 : $\widehat{\phi}_{1,2}^{(t)}$

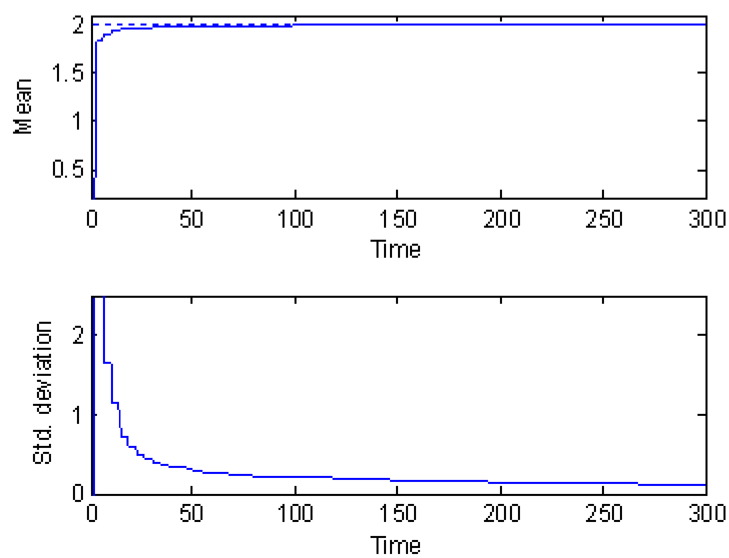


Figure 1.3 : $\widehat{\phi}_{1,3}^{(t)}$

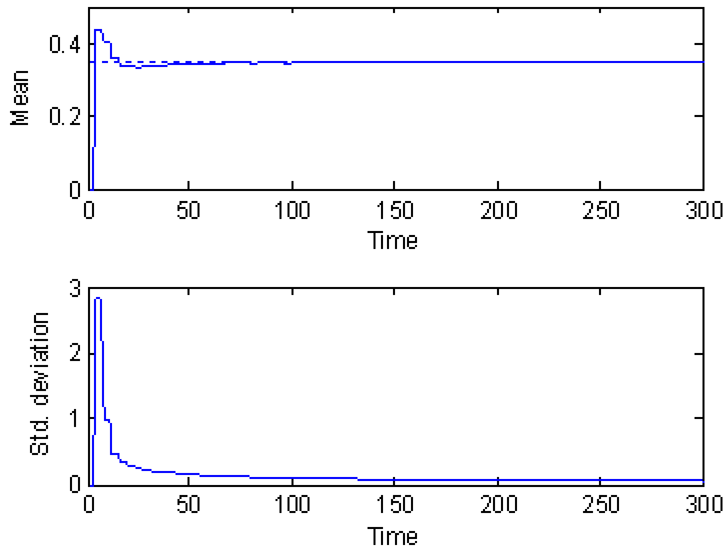


Figure 1.4 : $\widehat{\phi}_{1,4}^{(t)}$

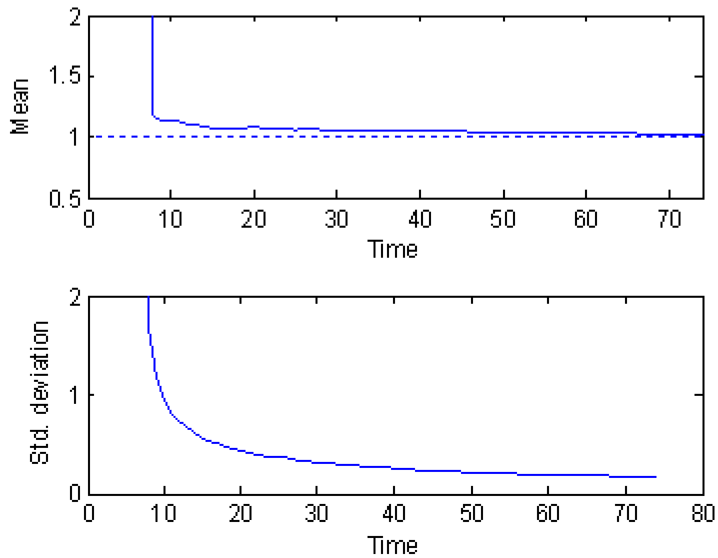


Figure 2.1 : $\widehat{\sigma}_1^{2(1,\tau)}$

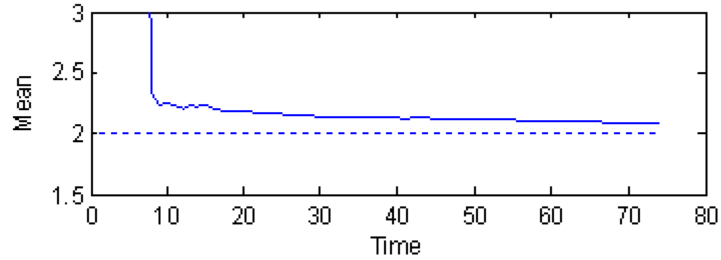


Figure 2.2 : $\widehat{\sigma}_2^{2(2,\tau)}$

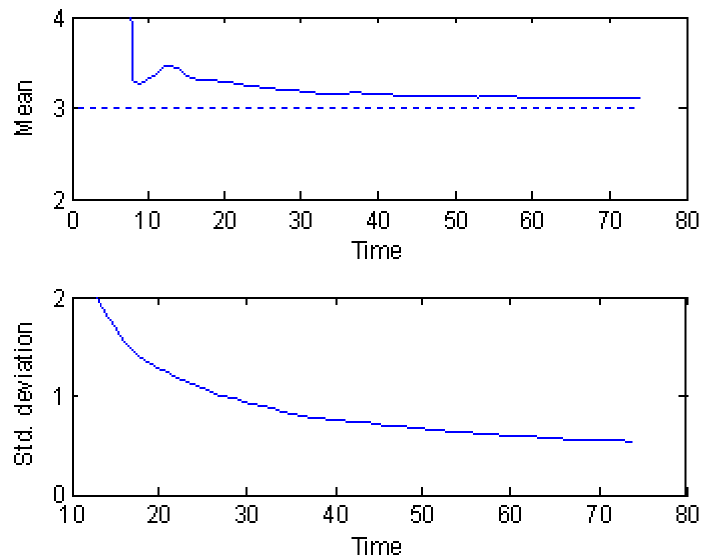


Figure 2.3 : $\widehat{\sigma}_3^{2(3,\tau)}$

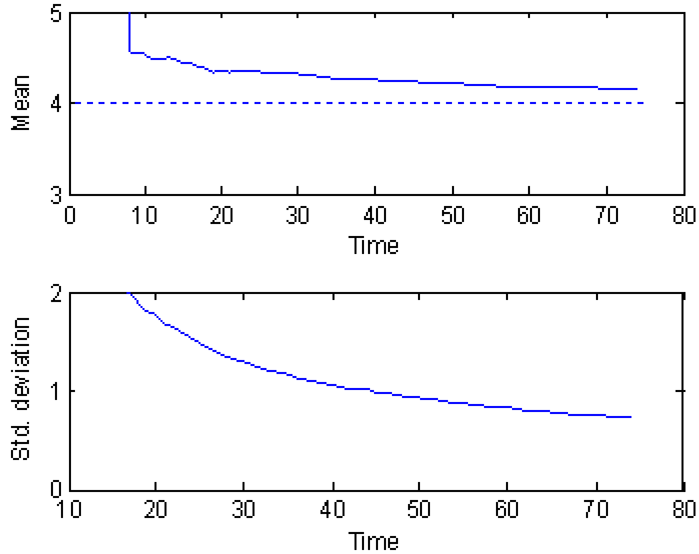


Figure 2.4 : $\widehat{\sigma}_4^{2(4,\tau)}$

3.8.2 Study of the $PRML$ algorithm

The $PRML$ algorithm (Algorithm 3.5.1) has been applied to $PARMA$ models of different orders ($PARMA_2(1,1)$, $PARMA_2(2,1)$) and for a variety of sample sizes (T). For each model, we have considered one thousand Monte Carlo replications. The unobservable starting values $\underline{y}'_{1,0} = (y_0, y_{-1}, ..., y_{1-p})'$ of the $PARMA$ stochastic difference equation and the initial parameter estimates are set as zero. The parameters and the variances for each season are given, respectively, by Table $C1.1.2$, and Table $C2.1.2$. The mean and standard deviation for the 1000 replications of the obtained estimates are reported in the same tables. In addition, Figure 3.s, $s = 1, 2$, illustrates the behavior of the moving average parameter means for the $PMA_2(1)$ model given by Table $D1.2$. We observe that the parameters are well estimated, but less accurate than the $PRLS$ algorithm. This is due to the fact that the $PRML$ algorithm optimizes an approximate least squares criterion (unlike the $PRLS$) when the model includes a moving average part.

Parameters estimation of a PARMA$_2$(1,1) model via the $PRML$ algorithm

Sample size / Parameter values	50	100	150	200
$\phi_{1,1} = 0.5$	0.5328 (0.1060)	0.5181 (0.0752)	0.5165 (0.0545)	0.5102 (0.0480)
$\theta_{1,1} = 0.9$	0.7972 (0.2090)	0.8485 (0.1452)	0.8716 (0.1076)	0.8761 (0.0885)
$\sigma_1^2 = 1$	0.9662 (0.3029)	0.9924 (0.2025)	0.9823 (0.1682)	0.9936 (0.1447)
$\phi_{2,1} = 1.4$	1.3767 (0.1749)	1.3794 (0.1085)	1.3856 (0.0931)	1.3867 (0.0797)
$\theta_{2,1} = -0.7$	-0.5638 (0.2393)	-0.6362 (0.1582)	-0.6413 (0.1246)	-0.6647 (0.1017)
$\sigma_2^2 = 1$	0.9883 (0.3045)	0.9967 (0.2116)	0.9939 (0.1633)	0.9906 (0.1470)

Table $C1.1.2$

Parameters estimation of a PARMA$_2$(2,1) model via the $PRML$ algorithm

Sample size / Parameter values	50	100	150	200
$\phi_{1,1} = 0.7$	0.7211 (0.3881)	0.7073 (0.2074)	0.7103 (0.1657)	0.6917 (0.1360)
$\phi_{1,2} = -0.5$	-0.5338 (0.1579)	-0.5078 (0.1118)	-0.5060 (0.0849)	-0.5044 (0.0716)
$\theta_{1,1} = -0.5$	-0.4763 (0.4198)	-0.4893 (0.2323)	-0.4924 (0.1864)	-0.5140 (0.1546)
$\sigma_1^2 = 1$	0.9408 (0.3087)	0.9525 (0.2040)	0.9659 (0.1668)	0.9848 (0.1408)
$\phi_{2,1} = 0.4$	0.3429 (0.2268)	0.3810 (0.1587)	0.3915 (0.1354)	0.3861 (0.1091)
$\phi_{2,2} = 0.3$	0.3279 (0.2290)	0.3137 (0.1608)	0.3033 (0.1346)	0.3095 (0.1067)
$\theta_{2,1} = 0.7$	0.5570 (0.3007)	0.6331 (0.2058)	0.6548 (0.1672)	0.6603 (0.1467)
$\sigma_2^2 = 1$	0.9154 (0.2960)	0.9519 (0.1920)	0.9707 (0.1629)	0.9694 (0.1399)

Table $C2.1.2$

$\theta \diagdown s$	1	2
θ_s	.6	.4
σ_s^2	1	1

Table $C0.1.2$

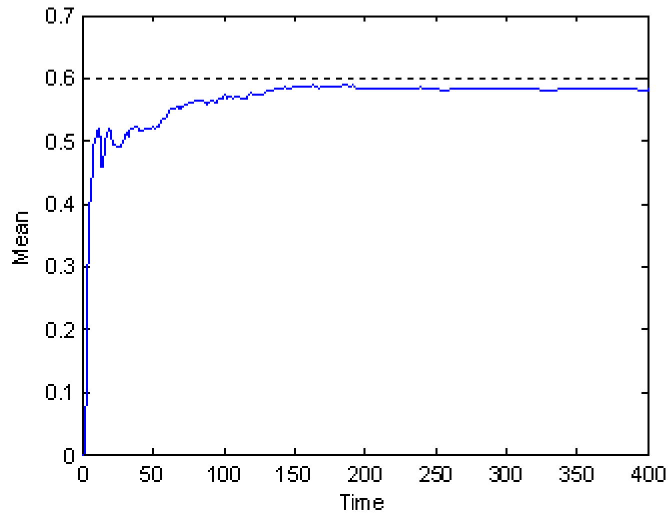


Figure 3.1 : $\widehat{\theta}_{1,1}^{(t)}$

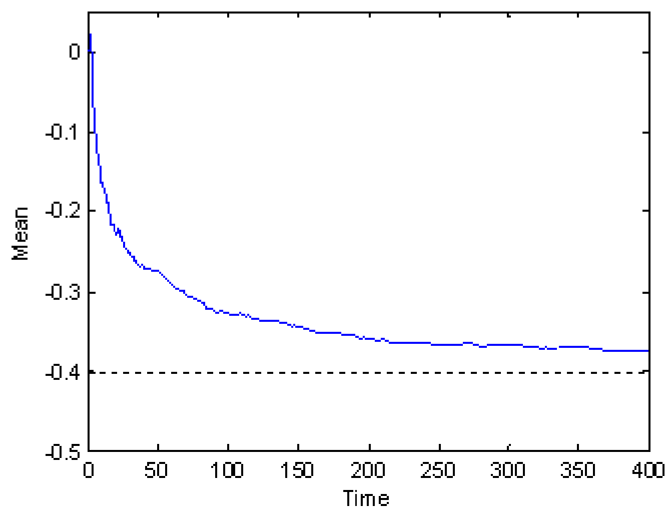


Figure 3.2 : $\widehat{\theta}_{2,1}^{(t)}$

3.9 Conclusion

Despite their remarkable importance, on-line estimation methods for $PARMA$ models remain relatively unexplored, especially when compared to their stationary $ARMA$ counterparts. This chapter has tried to close this gap by studying the on-line estimation problem for periodic time-varying models. We have proposed several algorithms for estimating $PARMA$ parameters by means of two different approaches, namely the pseudo-linear regression and recursive predictor error methods. For the particular PAR case, these two methods coincide with the recursive least squares, $PRLS$. The proposed methods parallel the ones available for standard time-invariant models. It has been shown that various desirable properties of the obtained algorithms take place in our periodic case.

Part II

Nonlinear modelling of periodic time series

Chapter 4

On periodic conditionally heteroskedastic models

4.1 Introduction

The class of periodic $GARCH$ models ($PGARCH$) introduced by Bollerslev and Ghysels (1996) have shown their appropriateness in capturing some interesting features of financial time series that cannot be explained by $PARMA$ or $GARCH$ models. Compared to the $PARMA$ modelling approach (Bessembinder and Hertzel, 1994; Abraham and Akenbery, 1993) for modelling economic time series with periodic dynamics, the appropriateness of $PGARCH$ models is partially explained and practically justified by their use of the stochastic conditional variance to capture the instantaneous volatility features and to allow the underlying coefficients to be periodic time-varying in order to capture the periodicity in the covariance structure of the generator process. However, this class of models captures only the periodicity in the conditional variance but not in the conditional mean. Franses and Paap (2000) have combined the PAR model for the conditional mean with the $PGARCH$ model for volatility. The resulting model has been called periodic autoregression (PAR) with $PGARCH(1,1)$ errors ($PAR-PGARCH$). They have investigated parameter estimation and tests based on the work of Bollerslev and Ghysels (1996), and applied with success their $PAR-PGARCH$ modelling to real daily data. However, for the asymptotic properties of the "inferenced" parameters, periodic stationarity and the existence of higher-order moments are required, while conditions under which these requirements are satisfied have not been found.

It is recognized that the existence of finite-order moments for a time series model is primordial for many statistical inference procedures, particularly for those dealing with the $ARCH$ type models, which require the existence of moments of higher orders than for the traditional linear models. In spite of this primary importance and interest of conditions, which assure the existence of higher-order moments for $PGARCH_S(q,\ p)$ models, it seems that there is no result tied to this problem. However, this problem was completely resolved for the classical $GARCH(q,\ p)$ models. Indeed, a necessary and sufficient condition for the existence of higher-order moments has been established for the general $ARCH(\ p)$ by Milhoj (1985), for the particular $GARCH(1,\ 1)$ by Bollerslev (1986) and for the general $GARCH(q,\ p)$ by Ling and McAleer (2002).

In this chapter, some probabilistic and algebraic properties of periodic $GARCH$ processes are investigated. Section 2 is concerned with the study of this class of models. In particular, we review the basic definitions, concepts and assumptions concerning this class. Moreover, the causality and invertibility properties of such models are studied in the first subsection. In the second one, we explore the existence of strict periodic stationary solutions of a $PGARCH$ difference equation. Indeed, we derive a periodic Markovian state space representation which is fundamental for the rest of this chapter and from which we obtain a necessary and sufficient periodic stationarity condition for a $PGARCH$ model. In the third subsection, we obtain, under general and tractable assumptions, a necessary and sufficient condition for general a $PGARCH_S(q,p)$ model to have finite higher-order moments. Our established condition extends the sufficient moment condition for $PGARCH$ models obtained by Bibi (2006) and coincides, for the particular case $S=1$ (non periodic) with the necessary and sufficient condition provided by Ling and McAleer (2002) for the classical $GARCH(p,q)$. In addition, an equivalent simple necessary and sufficient moment condition for the particular and useful first order $PGARCH(1,1)$ model is given. Section 3 studies the class of periodic autoregressions with period $ARCH$ errors $(PAR-PARCH)$. We propose a random coefficient formulation of the $PARCH$ part in order to describe parsimoniously the varying conditional variance feature. We call the new model *double periodic AR*. The existence of a strictly periodic stationary solution and the existence of higher-order moments for such a model are studied.

4.2 Periodic $GARCH$ models

We recall that a second order process $\{\varepsilon_t;\ t\in\mathbb{Z}\}$ is said to be a periodic autoregressive conditional heteroskedastic process with orders q and p and period S, denoted by $PGARCH_S(q,\ p)$, if it is given by

$$\begin{cases} \varepsilon_t=\sqrt{h_t}\zeta_t, \\ E\left(\varepsilon_t^2/\mathcal{F}_{t-1}\right)=h_t=\alpha_{t,0}\ +\ \sum\limits_{i=1}^{q}\alpha_{t,i}\,\varepsilon_{t-i}^2+\sum\limits_{j=1}^{p}\beta_{t,j}\,h_{t-j},\ t\in\mathbb{Z}, \end{cases} \tag{4.2.1}$$

where $\{\zeta_t,\ t \in \mathbb{Z}\}$ is an independent and identically distributed (*i.i.d.*) sequence of zero mean and unit variance. The symbol $\mathcal{F}_t$ denotes, as usual, the $\sigma-$algebra representing the available information up to time t. The parameters $\alpha_{t,i}$, and $\beta_{t,j}$ are periodic time varying with period S, i.e., $\alpha_{t+kS,i} = \alpha_{t,i}$ and $\beta_{t+kS,j} = \beta_{t,j}$, such that $\alpha_{t,0} > 0$, $\alpha_{t,i} \geq 0$, $\beta_{t,j} \geq 0$, $i = 1, ..., q$, $j = 1, ..., p$ and $\forall t \in \mathbb{Z}$.

Defining the innovation process $\{v_t;\ t \in \mathbb{Z}\}$, where $v_t = \varepsilon_t^2 - h_t = h_t\eta_t$ and $\eta_t = \zeta_t^2 - 1$, one can rewrite equations (4.2.1) as a $PARMA$ form for the squared process $\left\{\varepsilon_t^2;\ t \in \mathbb{Z}\right\}$ and $\{v_t;\ t \in \mathbb{Z}\}$

$$\left(1 - \phi_t\left(L\right)\right)\ \varepsilon_t^2 = \alpha_{t,0} + \left(1 - \beta_t\left(L\right)\right)\ v_t,\ t \in \mathbb{Z}, \tag{4.2.2}$$

where $\phi_t\left(L\right) = \sum\limits_{i=1}^{r} \phi_{t,i}\, L^i$, $\beta_t\left(L\right) = \sum\limits_{j=1}^{p} \beta_{t,j}\, L^j$ and $\phi_{t,i} = \alpha_{t,i} + \beta_{t,i}$, for $i = 1, ..., r = max\left\{p, q\right\}$ such that $\alpha_{t,i} = 0$ if $i > q$ and $\beta_{t,j} = 0$ if $j > p$. The symbol L denotes the backward shift operator, i.e., $L^j\, X_t = X_{t-j}$.

Of particular interest is the periodic $ARCH$ ($PARCH$) model obtained from model (4.2.1) when $\beta_{t,j} = 0$, $\forall t, j$. Thus a second order process $\{\varepsilon_t;\ t \in \mathbb{Z}\}$ is said to have a $PARCH$ representation with orders q and period S, denoted by $PARCH_S\left(q\right)$, if it is given by

$$\begin{cases} \varepsilon_t = \sqrt{h_t}\zeta_t, \\ E\left(\varepsilon_t^2 / \mathcal{F}_{t-1}\right) = h_t = \alpha_{t,0}\ +\ \sum\limits_{i=1}^{q} \alpha_{t,i}\, \varepsilon_{t-i}^2,\ t \in \mathbb{Z}. \end{cases} \tag{4.2.3}$$

4.2.1 Causality and Invertibility Conditions

Using the *period span lumping* based on the Gladyshev (1961) theorem, Bollerslev and Ghysels (1996) have obtained necessary and sufficient periodic stationarity conditions for the particular $PGARCH_2(1, 1)$ model. The establishment of an explicit result for a higher-order model using this approach, seems very difficult. However, using the so-called "*order span lumping*" introduced by Bentarzi and Hallin (1994) (cf, Chapter 1), we can derive explicit expressions of the causality and the invertibility for a general $PGARCH_S\left(q,\ p\right)$ model.

Periodically stationary condition

To determine a periodically stationary condition for a general $PGARCH_S\,(r,\,p)$ model given by (4.2.2), which is equivalent to model (4.2.1), we need to define, for $T \in \mathbb{Z}$, the $r-$variate processes $\underline{\epsilon}_T^2 = (\varepsilon_{Tr}^2, \varepsilon_{Tr-1}^2, ..., \varepsilon_{Tr-r+1}^2)'$ and $\underline{w}_T = (w_{Tr}, w_{Tr-1}, ..., w_{Tr-r+1})'$ where

$$w_{Tr-j} = \alpha_{Tr-j,0} + \left(1 - \beta_{Tr-j}\,(L)\right) v_{Tr-j}, \quad j = 0, ..., r-1,$$

and the following periodic matrices

$$\mathbf{\Phi}_{T,0} = \begin{pmatrix} 1 & -\phi_{Tr,1} & -\phi_{Tr,2} & \cdots & -\phi_{Tr,r-1} \\ 0 & 1 & -\phi_{Tr-1,1} & \cdots & -\phi_{Tr-1,r-2} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ & & & 1 & -\phi_{Tr-r+2,1} \\ 0 & \cdots & & 0 & 1 \end{pmatrix},$$

and

$$\mathbf{\Phi}_{T,1} = \begin{pmatrix} \phi_{Tr,r} & 0 & 0 & \cdots & 0 \\ \phi_{Tr-1,r-1} & \phi_{Tr-1,r} & 0 & \cdots & \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \phi_{Tr-r+2,2} & \phi_{Tr-r+2,3} & & \phi_{Tr-r+2,r} & 0 \\ \phi_{Tr-r+1,1} & \phi_{Tr-r+1,2} & & \phi_{Tr-r+1,r-1} & \phi_{Tr-r+1,r} \end{pmatrix}. \tag{4.2.4}$$

Using the previous notations, one can rewrite model (4.2.2) in the form

$$\mathbf{\Phi}_{T,0}\underline{\epsilon}_T^2 + \mathbf{\Phi}_{T,1}\underline{\epsilon}_{T-1}^2 = \underline{w}_T,\ T \in \mathbb{Z}, \tag{4.2.5}$$

from which we can obtain the following result.

Proposition 4.2.1

The $PGARCH_S(r,p)$ model given by (4.2.2) is second order periodically stationary (causal) if, and only if, the eigenvalues λ_i, $i = 1, ..., r$ of the positive definite matrix

$$\mathbf{Q} = \mathbf{\Phi}_{S,0}^{-1}\mathbf{\Phi}_{S,1}\mathbf{\Phi}_{S-1,0}^{-1}\mathbf{\Phi}_{S-1,1}\cdots\mathbf{\Phi}_{1,0}^{-1}\mathbf{\Phi}_{1,1},$$

are such that

$$0 \leq \lambda_i < 1.$$

Proof

We recall that the unique solution, on $\underline{\epsilon}_T^2$, of the first-order linear stochastic difference equation (4.2.5) is uniquely determined by giving one initial value $\underline{\epsilon}_s^2$, where $s \in \mathbb{Z}$ is the starting time of the process. Indeed, this solution is, for $T \geq s$, given by :

$$\underline{\epsilon}_T^2 = \sum_{u=s}^{T} G\left(u,s\right) \underline{w}_u + G\left(T,s\right)\ \underline{\epsilon}_s^2,$$

where $G\left(T,s\right)$ is the Green matrix associated to the periodic $r-$dimensional linear matrix difference operator of order 1,

$$\mathbf{\Phi}_T\left(L\right) = \mathbf{\Phi}_{T,0} + \mathbf{\Phi}_{T,1} L,$$

and which is the unique solution of the homogeneous matrix difference equation

$$\mathbf{\Phi}\left(L\right)\ G_T = \mathbf{0},$$

taking on initial value $\mathbf{\Phi}_{s,0}^{-1}$. It is easy to verify that the Green matrix is, in our case, given by

$$G\left(T,s\right) = \left(-1\right)^{T-s}\ \mathbf{\Phi}_{T,0}^{-1}\mathbf{\Phi}_{T,1}\ \mathbf{\Phi}_{T-1,0}^{-1}\mathbf{\Phi}_{T-1,1}\ ...\ \mathbf{\Phi}_{s+1,0}^{-1}\mathbf{\Phi}_{s+1,1}\ \mathbf{\Phi}_{s,0}^{-1}\ ,\quad T \geq s.$$

Letting $T - s = i + S\tau$ and $\Psi = \mathbf{\Phi}_{s+i,0}^{-1}\mathbf{\Phi}_{s+i,1}\mathbf{\Phi}_{s+i-1,0}^{-1}\mathbf{\Phi}_{s+i-1,1}...\mathbf{\Phi}_{s+i-(S-1),0}^{-1}\mathbf{\Phi}_{s+i-(S-1),1}$ one can rewrite the latter expression in the form

$$G\left(T,s\right) = \left(-1\right)^{T-s}\ \Psi^{\tau}\ \left(\mathbf{\Phi}_{s+i,0}^{-1}\mathbf{\Phi}_{s+i,1}\ \mathbf{\Phi}_{s+i-1,0}^{-1}\mathbf{\Phi}_{s+i-1,1}\ ...\ \mathbf{\Phi}_{s+1,0}^{-1}\mathbf{\Phi}_{s+1,1}\ \mathbf{\Phi}_{s,0}^{-1}\right).$$

Since the Euclidean norm of the matrix $\mathbf{\Phi}_{s+i,0}^{-1}\mathbf{\Phi}_{s+i,1}\ \mathbf{\Phi}_{s+i-1,0}^{-1}\mathbf{\Phi}_{s+i-1,1}\ ...\ \mathbf{\Phi}_{s+1,0}^{-1}\mathbf{\Phi}_{s+1,1}\ \mathbf{\Phi}_{s,0}^{-1}$ is uniformly bounded by $\max\limits_{0\leq i<S} \parallel \mathbf{\Phi}_{S,0}^{-1}\mathbf{\Phi}_{S,1}\ \mathbf{\Phi}_{S-1,0}^{-1}\mathbf{\Phi}_{S-1,1}\ ...\ \mathbf{\Phi}_{1,0}^{-1}\mathbf{\Phi}_{1,1}\ \mathbf{\Phi}_{s,0}^{-1} \parallel$ then a necessary and sufficient condition for the causality of the periodic operator $\mathbf{\Phi}_T\left(L\right)$ is given by

$$\lim_{\tau\rightarrow\infty} \left\|\Psi\right\|^{\tau} = 0,$$

and which accordingly holds if and only if the eigenvalues of the matrix Ψ (from where we obtain by a cyclic rearrangement the matrix $\mathbf{Q}$) are less than one (for further details about this approach one can see Bentarzi and Hallin, 1994; Ula and Smadi, 1997 and Bentarzi, 1998). ■

For the particular first-order $PGARCH_S(1,1)$ model, we have the corollary.

Corollary 4.2.2

The particularly simple $PGARCH_S(1,1)$ *model corresponding to* $q = p = 1$ *in* (4.2.2), *is second order periodically stationary if, and only if*

$$0 \leq \prod_{s=1}^{S}\left(\alpha_{s,1} + \beta_{s,1}\right) < 1.$$

Proof

For $p = q = 1$, we have $Q = \prod_{s=1}^{S} \left(\alpha_{s,1} + \beta_{s,1}\right)$. ■

It is worth noting that this condition, for the case $S = 2$, is the same causality condition obtained by Bollerslev and Ghysels (1996) for the $PGARCH_2(1,1)$ model with period 2.

It is of interest to rewrite Proposition 4.2.1 for the $PARCH$ model (4.2.3). Therefore we have the following corollary.

Corollary 4.2.3

The particular $PARCH_S(q)$ *model corresponding to* $p = 0$ *in* (4.2.2), *is second order periodically stationary if, and only if the eigenvalues* τ_i, $i = 1, ..., r$ *of the positive definite matrix*

$$\mathbf{R} = \boldsymbol{\alpha}_{S,0}^{-1}\boldsymbol{\alpha}_{S,1}\boldsymbol{\alpha}_{S-1,0}^{-1}\boldsymbol{\alpha}_{S-1,1}\cdots\boldsymbol{\alpha}_{1,0}^{-1}\boldsymbol{\alpha}_{1,1},$$

are such that

$$0 \leq \tau_i < 1, \ .$$

where $\alpha_{T,0}$ *and* $\alpha_{T,1}$ *are just* $\Phi_{T,0}$ *and* $\Phi_{T,1}$, *respectively, except that the coefficients* $\phi_{t,j}$ *in* (4.2.4) *are replaced by* $\alpha_{t,j}$.

Invertibility condition

Similarly to the last subsection, to determine an invertibility condition for the general

$PGARCH_S\,(r,\, p)$ model given by (4.2.2), we need to define, for $T \in \mathbb{Z}$, the $p-$variate processes $\underline{v}_T = (v_{Tp}, v_{Tp-1},\, ..., v_{Tp-p+1})'$ and $\underline{w}_T = (w_{Tp}, w_{Tp-1}, ..., w_{Tp-p+1})'$ where

$$w_{Tp-j} = -\alpha_{Tp-j,0} + \left(1 - \phi_{Tp-j}\left(L\right)\right)\, \varepsilon_{Tp-j}^2, \quad j = 0, ..., p-1,$$

and the following periodic matrices

$$\boldsymbol{\beta}_{T,0} = \begin{pmatrix} 1 & -\beta_{Tp,1} & -\beta_{Tp,2} & \cdots & -\beta_{Tp,p-1} \\ 0 & 1 & -\beta_{Tp-1,1} & \cdots & -\beta_{Tp-1,p-2} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ & & & 1 & -\beta_{Tp-p+2,1} \\ 0 & \cdots & & 0 & 1 \end{pmatrix},$$

and

$$\boldsymbol{\beta}_{T,1} = \begin{pmatrix} \beta_{Tp,p} & 0 & 0 & \cdots & 0 \\ \beta_{Tp-1,p-1} & \beta_{Tp-1,p} & 0 & \cdots & \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \beta_{Tp-p+2,2} & \beta_{Tp-p+2,3} & & \beta_{Tp-p+2,p} & 0 \\ \beta_{Tp-p+1,1} & \beta_{Tp-p+1,2} & & \beta_{Tp-p+1,p-1} & \beta_{Tp-p+1,p} \end{pmatrix}.$$

Using the previous notations, one can rewrite the model (4.2.2) in the form :

$$\boldsymbol{\beta}_{T,0}\underline{v}_T + \boldsymbol{\beta}_{T,1}\underline{v}_{T-1} = \underline{w}_T, \ T \in \mathbb{Z}. \tag{4.2.6}$$

From this form we can easily obtain the following result.

Proposition 4.2.3

The $PGARCH_S(q,p)$ model given by (4.2.2) is invertible if and only if the the eigenvalues μ_i, $i = 1, ..., p$, of the positive definite matrix

$$\Omega = \boldsymbol{\beta}_{S,0}^{-1}\boldsymbol{\beta}_{S,1}\boldsymbol{\beta}_{S-1,0}^{-1}\boldsymbol{\beta}_{S-1,1}\cdots\boldsymbol{\beta}_{1,0}^{-1}\boldsymbol{\beta}_{1,1}$$

are such that

$$0 \leq \mu_i < 1 \ . \tag{2.4.7}$$

Proof

The proof is based on the same mathematical techniques used to prove Proposition 4.2.1, hence it will be omitted.

4.2.2 Strict periodic stationarity

Letting $y_t = \varepsilon_t^2$ in (4.2.1), we obtain the following random coefficients representation

$$y_t = \alpha_{t,0}\zeta_t^2 + \sum_{i=1}^{q}\alpha_{t,i}\zeta_t^2 y_{t-i} + \sum_{j=1}^{p}\beta_{t,j}\zeta_t^2 h_{t-j}, \ t \in \mathbb{Z}, \tag{4.2.8}$$

which is more adequate for several theoretical purposes. Thoughout the rest of this chapter the vectorial relation $A > B$, means that each element of A is greater than the corresponding element of B.

Define the $(p+q)-$ random vectors

$$\begin{aligned} Y_t &= (y_t, ..., y_{t-q+1}, h_t, ..., h_{t-p+1})', \\ &\text{and} \\ B_t &= \left(\alpha_{t,0}\zeta_t^2, 0'_{(q-1)\times 1}, \alpha_{t,0}, 0'_{(p-1)\times 1}\right)', \end{aligned}$$

and the $(p+q)\times(p+q)$ random matrix A_t given by

$$A_t = \begin{pmatrix} \alpha_{t,1} \;\; \alpha_{t,2} \;\; \ldots \;\; \alpha_{t,q-1} \;\; \alpha_{t,q} & \beta_{t,1} \;\; \beta_{t,2} \;\; \ldots \;\; \beta_{t,p-1} \;\; \beta_{t,p} \\ I_{(q-1)\times(q-1)} \quad 0_{(q-1)\times 1} & 0_{(q-1)\times p} \\ \alpha_{t,1}\zeta_t^2 \;\; \alpha_{t,2}\zeta_t^2 \;\; \ldots \;\; \alpha_{t,q-1}\zeta_t^2 \;\; \alpha_{t,q}\zeta_t^2 & \beta_{t,1}\zeta_t^2 \;\; \beta_{t,2}\zeta_t^2 \;\; \ldots \;\; \beta_{t,p-1}\zeta_t^2 \;\; \beta_{t,p}\zeta_t^2 \\ 0_{(p-1)\times q} & I_{(p-1)\times(p-1)} \quad 0_{(p-1)\times 1} \end{pmatrix},$$

where $0_{m\times n}$ is the null matrix of order $m \times n$.

Then, one can rewrite the model (4.2.8) in the following Markovian state space representation

$$Y_t = A_t Y_{t-1} + B_t \tag{4.2.9}$$

in which the matrix A_t and the colon vector B_t are both stochastic, where their expectations are periodic in time with the same period and are given, respectively, by

$$\overline{A}_t = E(A_t) = \begin{pmatrix} \alpha_{t,1} \;\; \alpha_{t,2} \;\; \ldots \;\; \alpha_{t,q-1} \;\; \alpha_{t,q} & \beta_{t,1} \;\; \beta_{t,2} \;\; \ldots \;\; \beta_{t,p-1} \;\; \beta_{t,p} \\ I_{(q-1)\times(q-1)} \quad 0_{(q-1)\times 1} & 0_{(q-1)\times p} \\ \alpha_{t,1} \;\; \alpha_{t,2} \;\; \ldots \;\; \alpha_{t,q-1} \;\; \alpha_{t,q} & \beta_{t,1} \;\; \beta_{t,2} \;\; \ldots \;\; \beta_{t,p-1} \;\; \beta_{t,p} \\ 0_{(p-1)\times q} & I_{(p-1)\times(p-1)} \quad 0_{(p-1)\times 1} \end{pmatrix},$$

and

$$b_t = E(B_t) = (\alpha_{t,0}, 0'_{(q-1)\times 1}, \alpha_{t,0}, 0'_{(p-1)\times 1})'.$$

This Markovian state space representation was explored by Tong (1990) for $ARCH$ models, Chen and An (1998) and Ling and McAleer (2002) for studying the $GARCH$ model with constant coefficients and by Ling (1999*b*) in investigating the $CHARMA$ model. The recent review by Li et al. (2002) contains extensive literature on this subject. It is easy to verify that the unique solution of the stochastic difference equation (4.2.9), corresponding to the initial matrix value Y_s, where s

is the starting time, is given by :

$$Y_t = B_t + A_t A_{t-1} ... \ A_{s+2} A_{s+1} Y_s + \sum_{j=1}^{t-s-1} \left(A_t A_{t-1} ... A_{t-j+1} \right) B_{t-j} \tag{4.2.10}$$

The following proposition provides a necessary and sufficient condition for the existence of a periodically stationary solution of (4.2.10) as $s \to -\infty$.

Proposition 4.2.4

Suppose the $PGARCH$ process given by (4.2.1) *has finite second order moments. Then this process admits a strict periodically stationary solution of the form*

$$Y_t = B_t + \sum_{j=1}^{\infty} \prod_{i=0}^{j-1} A_{t-i} B_{t-j}, \tag{4.2.11}$$

if, and only if,

$$\rho \left(\prod_{s=1}^{S} \overline{A}_{S-s+1} \right) < 1. \tag{4.2.12}$$

Moreover, this solution is unique.

Proof

$i)$ To prove the sufficiency part, we note first that the independence of A_{t-j} and B_{t-k} for $j \neq k$, allows one to have

$$E \left(\sum_{j=1}^{\infty} \prod_{i=0}^{j-1} A_{t-i} B_{t-j} \right) = \sum_{j=1}^{\infty} \prod_{i=0}^{j-1} \overline{A}_{t-i} b_{t-j}. \tag{4.2.13}$$

If we can show that the right member of (4.2.13) is finite, then from the positivity of $\overline{A}_t$ and b_t we get

$$\sum_{j=1}^{\infty} \prod_{i=0}^{j-1} A_{t-i} B_{t-j} < \infty \quad a.s. \tag{4.2.14}$$

Letting $t = n^{(t)} + S\tau^{(t)}$ and $j = k^{(j)} + Sh^{(j)}$ such that $n^{(t)}, k^{(j)} \in \{1, ..., S\}$, $\tau \in \mathbb{Z}$ and $h^{(j)} \in \mathbb{N}$, one can, taking into account the $S-$periodicity of $\overline{A}_t$ and b_t, rewrite the expectation in (4.2.13) as :

$$\sum_{j=1}^{\infty} \left(\prod_{l=1}^{S} \overline{A}_{n^{(t)}-l+1} \right)^{h^{(j)}} \left(\prod_{l=1}^{k^{(j)}} \overline{A}_{n^{(t)}-l+1} \right) b_m \tag{4.2.15}$$

where $h^{(j)} = [j/S]$ and $m \in \{1, ..., S\}$ is such that $n^{(t)} - k^{(j)} = m + Sn$ with $n \in \mathbb{Z}$. Since $\left(\prod_{l=1}^{k^{(j)}} \overline{A}_{n^{(t)}-l+1}\right)$ is uniformly bounded by $\max_{0 \le i \le S} \left\| \prod_{l=1}^{S} A_{i-l+1} \right\|$ then a sufficient condition for (4.2.14) to hold is given by

$$\lim_{j \to \infty} \left\| \prod_{l=1}^{S} \overline{A}_{n^{(t)}-l+1} \right\| = 0,$$

which accordingly holds if and only if the eigenvalues of the matrix $\prod_{l=1}^{S} \overline{A}_{n^{(t)}-l+1}$ (from where we obtain by a cyclic rearrangement the matrix $\prod_{s=1}^{S} \overline{A}_{S-s+1}$) are less than one. Hence, if (4.2.12) holds, one can easily verify that Y_t, given by (4.2.11), verifies (4.2.9).

ii) Necessity of (4.2.12). From (4.2.9), we have for some $s \in \{1, ..., S\}$

$$Y_s = B_s + \prod_{i=0}^{k-1} A_{s-i} Y_{s-k} + \sum_{j=1}^{k-1} \prod_{i=0}^{j-1} A_{s-i} B_{s-j},$$

and because of the independence of A_{t-j} and B_{t-i} for $j \neq i$, the non-negativity of Y_t, B_t and A_t and the finiteness of $E\left(Y_s\right)$ for all $s \in \{1, ..., S\}$, we have for some positive integer k, $(k \geq 1)$

$$E\left(Y_s\right) \geq c_1 \sum_{j=1}^{k-1} \prod_{i=0}^{j-1} \overline{A}_{s-i} B,$$

which, as $k \to \infty$, implies that

$$\sum_{j=1}^{\infty} \prod_{i=0}^{j-1} \overline{A}_{s-i} B < \infty, \tag{4.2.16}$$

where the constant c_1 is defined as the minimum of all positive elements of $E\left(B_s\right)$ for all $s = 1, ..., S$, and $B = (1, 0'_{(q-1)\times 1}, 1, 0'_{(p-1)\times 1})'$. Adopting the same notations as in (4.2.15) with s in place of t, one can rewrite (4.2.16) as

$$\sum_{j=1}^{\infty} \left(\prod_{l=1}^{S} \overline{A}_{s-l+1}\right)^{[j/S]} \left(\prod_{l=1}^{k^{(j)}} \overline{A}_{s-l+1}\right) B < \infty. \tag{4.2.17}$$

Note that from (4.2.17) we cannot claim directly that

$$\sum_{j=1}^{\infty}\left(\prod_{l=1}^{S}\overline{A}_{s-l+1}\right)^{[j/S]} < \infty. \tag{4.2.18}$$

However, because of the particular forms of A_t and B, which are the same as for the $GARCH$ case, and the nonnegativity of their elements, one can, similarly to Ling and McAleer (2002, formula (4.5) p. 727), show that

$$\left(\prod_{l=k^{(j)}-r+1}^{k^{(j)}}\overline{A}_{s-l+1}\right)B = \left(\prod_{l=k^{(j)}-r+1}^{0}\overline{A}_{s-l+1}\right)\left(\prod_{l=1}^{k^{(j)}}\overline{A}_{s-l+1}\right)B > 0, \tag{4.2.19}$$

for all $r \geq r^*$ where $r^* = \max(p, q)$. Thus, from (4.2.18) and (4.2.19), we have for some $r \geq r^*$

$$\sum_{j=1}^{\infty}\left(\prod_{l=1}^{S}\overline{A}_{s-l+1}\right)^{[j/S]}\left(\prod_{l=k^{(j)}-r+1}^{k^{(j)}}\overline{A}_{s-l+1}\right)B < \infty. \tag{4.2.20}$$

Let $\mathbf{r}_{ik}^{(j)}$ be the (i,k)-th element of $\left(\prod_{l=1}^{S}\overline{A}_{s-l+1}\right)^{[j/S]}$. From (4.2.20) one can deduce that $\sum_{j=1}^{\infty}\mathbf{r}_{ik}^{(j)} < \infty$ for all i,k. This implies (4.2.18) and hence (4.2.12). The proof of uniqueness is straightforward and hence omitted.

Remark 4.2.1

Setting $\widetilde{A}_{s+S\tau} = \prod_{l=1}^{S} A_{s+S\tau-l+1}$ and $\widetilde{B}_{s+S\tau} = \sum_{j=1}^{S}\prod_{l=1}^{S} A_{s+S\tau-l+1}B_{s-j}$, under the assumptions that $\{\widetilde{A}_{s+S\tau}\}_\tau$ and $\{\widetilde{B}_{s+S\tau}\}_\tau$ are stationary processes and $\{\widetilde{A}_{s+S\tau}\}_\tau$ has a negative Lyapunov exponent, we can use Brandt's theorem (see Brandt, 1986 and Bougerol and Picard, 1992*b*) to show that $\{Y_{s+S\tau}\}_\tau$ has a unique stationary solution. We recall that the sequence $\{Y_{s+S\tau}\}_\tau$ is the $PGARCH$ process sampled every S points and at lag s. We also note that a sufficient condition for $\{A_{s+S\tau}\}_\tau$ to have a negative Lyapunov exponent is that the largest absolute eigenvalue of $E(\widetilde{A}_{s,s+S\tau})$ is less than one (see for example Kesten and Spitzer, 1984, equation (1.4)).

4.2.3 Existence of finite moments for $PGARCH_S(q,p)$ models

For a classical $GARCH$ model, i.e. when $S=1$ in (4.2.1), Chen and An (1998) have stated that a sufficient condition for the existence of $E\left(\varepsilon_t^{2m}\right)$ is that $E\left(\zeta_t^{2(m-1)}\right)$ is finite and $\rho\left(E\left(A_t^{\otimes m}\right)\right)<1$. Ling (1999$a$) obtained the same condition by using another approach. More recently, Ling and McAleer (2002) have shown that only the last condition is sufficient and necessary. The following proposition generalizes Ling and McAleer (2002)'s result to the class of $PGARCH$ models.

Proposition 4.2.5

Let ε_t be a strictly periodically stationary solution of (4.2.1). *Then a necessary and sufficient condition for $E\left(\varepsilon_t^{2m}\right)$ to be finite is that*

$$\rho\left(\prod_{s=1}^{S}E\left(A_{S-s+1}^{\otimes m}\right)\right)<1. \tag{4.2.21}$$

Proof

i) The proof of the sufficiency part of the proposition is based on the weak convergence theory (see Billingsley, 1968) which stipulates that a sufficient condition for the existence of $E\left(\varphi(X)\right)$ is that

$$\lim_{\tau\to\infty}\inf E\left(\varphi(X_\tau)\right),$$

is finite, where (X_τ) is a sequence of random variables which converge in distribution to the random variable X and $\varphi(.)$ is a measurable function. Let $Y_{s+S\tau}$ be the solution of (4.2.9) corresponding to the initial value $Y_0=0$ and Y^* a random vector having the same distribution of (4.2.11) with $t=s$, for some $s\in\{1,...,S\}$. It is clear that the S-periodic feature of the process generated by (4.2.9), implies the convergence in distribution, as $\tau\to+\infty$, of the solution $Y_{s+S\tau}$ to Y^*. Otherwise, taking expectation on both sides of (4.2.9), with $t=s+S\tau$, we obtain

$$U_t-\overline{A}_tU_{t-1}=b_t, \tag{4.2.22}$$

where $U_t=E\left(Y_t\right)$. Equation (4.2.22) is a simplest S-periodic difference equation of order 1. In view of Proposition 4.2.4 (see also Bentarzi and Hallin, 1994; and Bentarzi, 1998), we see that $\lim\limits_{\tau\to\infty}U_{s+S\tau}$ exists for every $s\in\{1,...,S\}$ if (4.2.21) holds for $m=1$, in which case, $\{U_{s+S\tau}\}$ is bounded.

Now, define $V_t = E\left(vec\left(Y_t Y_t'\right)\right)$, then from (4.2.9), we get

$$V_t = E\left(A_t^{\otimes 2}\right) V_{t-1} + \left[E(A_t \otimes B_t) + E(B_t \otimes A_t)\right] U_t + vec(E\left(B_t B_t'\right)), \tag{4.2.23}$$

which is an ordinary S-periodic difference equation of order 1, where $E\left(A_t^{\otimes 2}\right)$, $E(A_t \otimes B_t)$ and $vec(E\left(B_t B_t'\right))$ are periodic finite matrices. Since the last two terms of the right side of (4.2.23) are bounded, then using again Proposition 4.2.4, we conclude that $\lim\limits_{\tau\to\infty} V_{s+S\tau}$ exists for every $s \in \{1, ..., S\}$ if (4.2.21) holds for $m = 2$. Finally because $y_{s+S\tau}^2$ converges to y_s^{*2} and $\lim\limits_{\tau\to\infty} V_{s+S\tau}$ is finite, then $E\left(y_s^{*2}\right) < \infty$ which completes the proof for $m = 2$. For a general m the proof is similar and hence omitted.

ii) It is clear that if $E\left(\varepsilon_t^{2m}\right) < \infty$, then $E\left(Y_t^{\otimes m}\right) < \infty$. Because of the non-negativity of Y_t, A_t and B_t, and the periodicity of their m-order moments, we have for some positive k,

$$\begin{aligned}
E\left(Y_t^{\otimes m}\right) &\geq E(A_t Y_{t-1})^{\otimes m} + E\left(B_t^{\otimes m}\right), \\
&\geq E(A_t^{\otimes m}) E(Y_{t-1}^{\otimes m}) + c_1 B^{\otimes m}, \\
&\geq c_2 \sum_{j=1}^{k} \left(\prod_{i=1}^{j} E(A_{t-i+1}^{\otimes m}) \right) B^{\otimes m}, \\
&= c_2 \sum_{j=1}^{k} \left(\prod_{l=n^{(t)}}^{n^{(t)}+S-1} E(A_l^{\otimes m}) \right)^{h^{(j)}} \left(\prod_{l=n^{(t)}}^{n^{(t)}+s^{(j)}-1} E(A_l^{\otimes m}) \right) B^{\otimes m},
\end{aligned} \tag{4.2.24}$$

where as in (4.2.15) $t = n^{(t)} + S\tau^{(t)}$ and $j = s^{(j)} + Sh^{(j)}$, for $j = 1, ..., k$ such that $n^{(t)}, s^{(j)} \in \{1, ..., S\}$, $\tau^{(t)} \in \mathbb{Z}$ and $h^{(j)} \in \mathbb{N}$. The constant c_2 is defined as the minimum of all positive elements of $E\left(B_s^{\otimes m}\right)$ for all $s = 1, ..., S$. Hence, if $k \to \infty$, then from (4.2.24) we obtain

$$\sum_{j=1}^{\infty} \left(\prod_{l=1}^{S} E(A_{n^{(t)}-l+1}^{\otimes m}) \right)^{[j/S]} \left(\prod_{l=1}^{s^{(j)}} E(A_{n^{(t)}-l+1}^{\otimes m}) \right) B^{\otimes m} < \infty, \tag{4.2.25}$$

from which we cannot obtain directly (4.2.21). However, similarly to Ling and McAleer (2002, p727), one can show

$$\left(\prod_{l=s^{(j)}-r+1}^{s^{(j)}} E(A_{n^{(t)}-l+1}^{\otimes m}) \right) B^{\otimes m} = \left(\prod_{l=s^{(j)}-r+1}^{0} E(A_{n^{(t)}-l+1}^{\otimes m}) \prod_{l=1}^{s^{(j)}} E(A_{n^{(t)}-l+1}^{\otimes m}) \right) B^{\otimes m} > 0, \tag{4.2.26}$$

for all $r \geq r^* = \max(p,q)$. Hence from (4.2.25) and (4.2.26) we get

$$\sum_{j=1}^{\infty}\left(\prod_{l=1}^{S}E(A_{n^{(t)}-l+1}^{\otimes m})\right)^{[j/S]}\left(\prod_{l=s^{(j)}-r+1}^{s^{(j)}}E(A_{n^{(t)}-l+1}^{\otimes m})\right)B^{\otimes m}<\infty,$$

which, by (4.2.26) implies that

$$\sum_{j=1}^{\infty}\left(\prod_{l=1}^{S}E(A_{n^{(t)}-l+1}^{\otimes m})\right)^{[j/S]}<\infty,$$

for all $n^{(t)} \in \{1,...,S\}$, from where we obtain by a cyclic rearrangement the result (4.2.21). ■

Remark 4.2.2

Similar results to the ones given by Proposition 4.2.4 and to the sufficiency part of Proposition 4.2.5 have been obtained in Bibi (2004). However, our proofs are conducted in a different way. Moreover, the necessity part of Proposition 4.2.5 seems not to have been explored hitherto.

4.2.4 Necessary and sufficient moment condition for the PGARCH$_S$(1,1) model

Suppose that for the particular $PGARCH_S(1,1)$ model, the moments $E\left(\zeta_t^{2m}\right)$ exist for any positive integer m and that the process starts infinitely far in the past with finite $2m$-th moment. Then we have the following result.

Proposition 4.2.6

A necessary and sufficient condition for the existence of the $2m$-order moment for the $PGARCH_S(1,1)$ process is that

$$\prod_{s=0}^{S-1}\left(\sum_{l=0}^{m}\binom{m}{l}\alpha_{S-s,1}^{m-l}\beta_{S-s,1}^{l}a_{m-l}\right)<1, \tag{4.2.27}$$

where $a_j = E\left(\zeta_t^{2j}\right)$.

Proof

For the $PGARCH_S(1,1)$ model the binomial formula yields

$$E(\varepsilon_t^{2m}) = h_t^m a_m = \left(\sum_{k=0}^{m}\binom{m}{k}\alpha_{t,0}^{m-k}\sum_{l=0}^{k}\binom{k}{l}\alpha_{t,1}^{k-l}\varepsilon_t^{2(k-l)}\beta_{t,1}^{l}h_{t-1}^{l}\right)a_m. \tag{4.2.28}$$

Since

$$E\left(\varepsilon_t^{2(k-l)} h_{t-1}^l \,/\mathcal{F}_{t-2}\right) = h_{t-1}^k a_{k-l},$$

we have

$$E\left(h_t^m \,/\mathcal{F}_{t-2}\right) = \sum_{k=0}^{m} \binom{m}{k} \alpha_{t,0}^{m-k} h_{t-1}^k \sum_{l=0}^{k} \binom{k}{l} \alpha_{t,1}^{k-l} \beta_{t,1}^l a_{k-l}. \tag{4.2.29}$$

Let $W_t = (h_t^m, h_t^{m-1}, ..., h_t)'$, then from (4.2.28) we know that $E(\varepsilon_t^{2m})$ is a linear combination of W_t and from (4.2.29) we find

$$E\left(W_t \,/\mathcal{F}_{t-2}\right) = b_t + A_t W_{t-1}, \tag{4.2.30}$$

where the S-periodic A_t is an upper triangular matrix where its diagonal elements are given by $\sum\limits_{l=0}^{k} \binom{k}{l} \alpha_{t,1}^{k-l} \beta_{t,1}^l a_{k-l}$. By successive substitution in (4.2.30), we obtain

$$E\left(W_t \,/\mathcal{F}_{t-l-1}\right) = b_t + \sum_{k=0}^{l} \prod_{j=1}^{k} A_{t-j+1} b_{t-j} + \left(\prod_{j=1}^{l+1} A_{t-j+1}\right) W_{t-l},$$

and because of the periodicity of A_t, and using Proposition 4.2.4 again, one can see that

$$\lim_{l \longrightarrow \infty} E\left(W_t / \mathcal{F}_{t-l-1}\right) = E\left(W_t\right),$$

exists if, and only if, the eigenvalues of the upper triangular matrix $\prod\limits_{s=0}^{S-1} A_{S-s}$ lie inside the unit circle, which straightforwardly implies (4.2.27). ■

Finally, note that condition (4.2.27) is the same for $S = 1$ as that given by Bollerslev (1986) for the classical $GARCH(1,1)$ case. Such a condition, although it requires the assumption that the process starts infinitely far in the past with finite $2m$-th moment, seems simpler than the condition (4.2.30) for the particular $PGARCH_S(1,1)$.

4.3 On double periodic AR models

The main goal of this section is to study some probabilistic properties of the class of S-periodic autoregressive models with constant order p and with periodic $ARCH(q)$ errors (denoted by $PAR-PARCH_S\,(p,q)$). We first propose a new formulation of the $PAR - PARCH_S\,(p,q)$ model in which the innovation equation has periodically correlated random coefficients, in order to describe parsimoniously the varying conditional variance feature. The proposed model that we call *double periodic AR* is, from a second order point of view, equivalent to the $PAR-PARCH$ model obtained in Franses and Paap (2000), in the sense that the two models have the same conditional expectations and variances. The simplicity of the proposed model allows us to investigate easily the strict-sense periodic stationarity property and the existence of higher-order moments. The remainder of this section is organized as follows. The first subsection introduces a random coefficient formulation of the $PAR-PARCH_S\,(p,q)$ and some suitable assumptions. In the second subsection we derive a periodic Markovian state space representation of a $PAR-PARCH_S\,(p,q)$ model, from which we obtain necessary and sufficient periodic stationarity conditions. Furthermore, we give a sufficient condition for a $PAR-PARCH_S\,(p,q)$ model to have finite higher-order moments.

4.3.1 Definitions and assumptions

In this section, we say that the second order process $\{y_t,\ t\in\mathbb{Z}\}$ has a periodic autoregressive representation with periodic conditionally heteroskedastic errors of order p and q and period S, denoted here by $PAR-PARCH_S\,(p,q)$, if it is a solution of stochastic difference equations of the form

$$\begin{cases} y_t = \sum\limits_{j=1}^{p} \phi_{tj} y_{t-j} + \varepsilon_t, \\ \varepsilon_t = \sqrt{h_t}\zeta_t, \\ h_t = E(\varepsilon_t^2 / \mathcal{F}_{t-1}) = \alpha_{t,0} + \sum\limits_{i=1}^{q} \alpha_{t,i}\varepsilon_{t-i}^2,\ t\in\mathbb{Z}, \end{cases} \tag{4.3.1}$$

where $\{\zeta_t,\ t\in\mathbb{Z}\}$ are independent and identically distributed (*i.i.d.*) with zero mean and unit variance and where $\mathcal{F}_t$ denotes the $\sigma-$algebra representing the information up to time t. The autoregressive and the heteroskedastic parameters ϕ_{tj} and $\alpha_{t,i}$ are periodic in time with period S, i.e., $\phi_{t+kS,j} = \phi_{t,j}$, $j = 1,...,p$ and $\alpha_{t+kS,i} = \alpha_{t,i}$, $i = 0,...,q$, such that $\alpha_{t,0} > 0$, $\alpha_{t,i} \geq 0$, $i = 1,...,q$, $\forall t\in\mathbb{Z}$.

Under some suitable assumptions given below (**A1-A5**), the model (4.3.1) can be represented equivalently, in a second order viewpoint, via the following double periodic AR model in which the innovation equation has periodically correlated random coefficients, that is

$$\begin{cases} y_t = \sum\limits_{j=1}^{p} \phi_{tj} y_{t-j} + \varepsilon_t & \text{(Observation equation)}, \\ \varepsilon_t = e_t + \sum\limits_{i=1}^{q} \delta_{t,i} \varepsilon_{t-i} & \text{(Innovation equation)}, \end{cases} \tag{4.3.2}$$

where the sequences of random variables $\{e_t\}$ $\{y_t\}$ $\{\varepsilon_t\}$ and random vectors $\{\delta_t\}$, with $\delta_t = (\delta_{t1}, \delta_{t2}, ..., \delta_{tq})'$, obey the following assumptions.

A1 : $\{e_t\}$ is a sequence of independent variables with zero mean and $S-$periodic variance $E\left(e_t^2\right) = \alpha_{t,0}$, where $\alpha_{t,0}$ is the same constant as the one given in (4.3.1);

A2 : $\{\delta_t\}$ is a sequence of independent vectors with zero mean and covariance matrix $\Sigma_t = diag\left(\alpha_{t,1}, \alpha_{t,2}, ..., \alpha_{t,q}\right)$, where $\alpha_{t,j}, j = 1, ..., q$ are defined in (4.3.1);

A3 : the sequences $\{e_t\}$ and $\{\delta_t\}$ are mutually independent;

A4 : $\{e_t\}$ and $\{\delta_t\}$ are S-periodically distributed, i.e. the distribution of δ_s is the same as that of $\delta_{s+S\tau}$ for all $s \in \{1, ..., S\}$, $\tau \in \mathbb{Z}$ and so for $\{e_t\}$;

A5 : y_t and ε_t are $\mathcal{F}_t-$measurable.

In **A2**, the assumption that the covariance matrix Σ_t is diagonal can be relaxed to a more general covariance matrix. This leads to a periodic augmented $ARCH$ formulation (denoted here by $P-AARCH$), which coincides for $S = 1$ with the augmented $ARCH$ model ($AARCH$) proposed by Bera et al. (1992). It is worth noting that with the conditional normality assumption, i.e. if $\{e_t\}$ and $\{\delta_t\}$ are jointly normal, and $\{\zeta_t, t \in \mathbb{Z}\}$ are normally distributed, models (4.3.1) and (4.3.2) are higher-order equivalent, in the sense that they have the same conditional higher-order moments. This condition is not required in the derivation of the probabilistic properties of model (4.3.2) and hence can be omitted here. The specification of the conditional heteroskedastic innovation equation as an autoregression with random coefficients was introduced first by Tsay (1987) for conditional heteroskedastic $ARMA$ models ($CHARMA$), generalized by Wong and Li (1997) to the multivariate case and investigated by Li et al. (2001) for a partially nonstationary AR model with conditional heteroskedasticity. A similar representation applied to $GARCH$ models have been proposed by Bera et al. (1996). All of these works assume the strict stationarity of the random coefficients, while in our case, the random coefficients are assumed to be periodically distributed. In the rest of this paper, we use only the second representation, which is very useful to derive some basic properties of the model, whereas the representation (4.3.1) can be used to investigate the parameters estimation problem.

4.3.2 Markovian representation of the model

In this subsection the $PAR-PARCH_S(p,q)$ model given by (4.3.2) will be written in a Markovian form which allows one to study easily the periodic stationarity, periodic ergodicity and existence of higher order moments of the model. We rewrite model (4.3.2) as follows

$$\begin{cases} y_t = \sum\limits_{j=1}^{p} \phi_{tj} y_{t-j} + \sum\limits_{i=1}^{q} \delta_{t,i}\varepsilon_{t-i} + e_t, \\ \varepsilon_t = \sum\limits_{i=1}^{q} \delta_{t,i}\varepsilon_{t-i} + e_t, \end{cases} \tag{4.3.3}$$

and define the $(p+q)-$random vectors $Y_t = (y_t, ..., y_{t-p+1}, \varepsilon_t, ..., \varepsilon_{t-q+1})'$ and

$B_t = (e_t, 0'_{(p-1)\times 1}, e_t, 0'_{(q-1)\times 1})'$, and the $(p+q)\times(p+q)$ random matrix A_t

$$A_t = \begin{pmatrix} \begin{matrix} \phi_{t,1} & \phi_{t,2} & \dots & \phi_{t,p-1} \\ & I_{(p-1)\times(p-1)} & & \end{matrix} & \begin{matrix} \phi_{t,p} \\ 0_{(p-1)\times 1} \end{matrix} & \begin{matrix} \delta_{t,1} & \delta_{t,2} & \dots & \delta_{t,q-1} & \delta_{t,q} \\ & & 0_{(p-1)\times q} & & \end{matrix} \\ 0_{q\times p} & & \begin{matrix} \delta_{t,1} & \delta_{t,2} & \dots & \delta_{t,q-1} & \delta_{t,q} \\ & I_{(q-1)\times(q-1)} & & & 0_{(q-1)\times 1} \end{matrix} \end{pmatrix},$$

where $0_{m\times 1}$ is the null vector of order m.

Using these notations, one can rewrite model (4.3.3) in the following equivalent Markovian representation :

$$Y_t = A_t Y_{t-1} + B_t, \tag{4.3.4}$$

in which the matrix A_t and the vector B_t are both stochastic with $S-$periodic higher order moments in view of assumption A4. This Markovian representation was explored first by Nicholls and Quinn (1982) for random coefficients autoregressive models (RCA), Tsay (1987) and Ling (1999) in investigating the $CHARMA$ model, Tong (1990) for studying $ARCH$ models, Ling and Deng (1993) for VAR models with conditional heteroskedastic errors and Bougerol and Picard (1992), Chen and An (1998) and Ling and McAleer (2002) for studying the $GARCH$ model with constant coefficients (see Li et al., 2002).

4.3.3 Strict periodic stationarity

It is easy to verify that the unique solution, for the starting time s, of the stochastic difference equation (3.2) is given by

$$Y_t = B_t + A_t A_{t-1} A_{t-2} \ldots A_{s+2} A_{s+1} Y_s + \sum_{j=1}^{t-s-1} \left(A_t A_{t-1} A_{t-2} \ldots A_{t-j+1}\right) B_{t-j}. \tag{4.3.5}$$

Furthermore, the following theorem gives a sufficient condition and another necessary condition for the existence of a unique strict periodically stationary solution of (4.3.5) as $s \to -\infty$.

Theorem 4.3.1 (Aknouche and Guerbyenne, 2005)

In order that the $PAR-PARCH_S(p,q)$ model (4.3.2) *admits a unique $\mathcal{F}_t-$measurable periodically stationary solution, in the strict sense, of the form*

$$Y_t = B_t + \sum_{j=1}^{\infty} \prod_{i=0}^{j-1} A_{t-i} B_{t-j}, \tag{4.3.6}$$

it is sufficient that

$$\rho\left(\prod_{s=0}^{S-1} E\left(A_{S-s}^{\otimes 2}\right)\right) < 1, \tag{4.3.7}$$

and necessary that

$$\lim_{j \longrightarrow \infty} \left\| \left(\prod_{l=0}^{S-1} E\left(A_{n^{(t)}-l}^{\otimes 2}\right)\right)^j B^{\otimes 2} \right\| = 0, \tag{4.3.8}$$

where $B = (1, 0'_{(p-1)\times 1}, 1, 0'_{(q-1)\times 1})'$.

Proof

i) Proof of sufficiency. Let $P_{t,j-1} = \prod_{i=0}^{j-1} A_{t-i}$ and $S_{t,m} = \sum_{j=0}^{m} P_{t,j-1} B_{t-j}$, where the notation $\prod_{i=x}^{y} A_i = 1$ for $x > y$ is used. Obviously we have

$$Y_t = S_{t,m} + R_{t,m}, \tag{4.3.9}$$

where $R_{t,m} = \prod_{i=0}^{m} A_{t-i} Y_{t-m-1}$. To prove the sufficiency part it is enough to show that $E\left(S_{t,m}S'_{t,m}\right)$ is finite as $m \longrightarrow \infty$, for all fixed $t \in \{1, ..., S\}$. From the independence of B_t and A_t and that of A_{t-i} and A_{t-j} for $j \neq i$, we have

$$\begin{aligned} vec\left[E\left(S_{t,m}S'_{t,m}\right)\right] &= vec\left[\sum_{j=0}^{m} E\left(P_{t,j-1}B_{t-j}B'_{t-j}P'_{t,j-1}\right)\right], \\ &= E\left[\sum_{j=0}^{m} P_{t,j-1}^{\otimes 2} vec\left(B_{t-j}B'_{t-j}\right)\right], \\ &= \left[\sum_{j=0}^{m}\prod_{i=0}^{j-1} E\left(A_{t-i}^{\otimes 2}\right) vecE\left(B_{t-j}B'_{t-j}\right)\right]. \end{aligned}$$

Letting $t = n^{(t)} + S\tau^{(t)}$ and $j = n^{(j)} + S\tau^{(j)}$ such that $n^{(t)}, n^{(j)} \in \{1, ..., S\}$, $\tau^{(t)} \in \mathbb{Z}$ and $\tau^{(j)} \in \mathbb{N}$, one can write the limit $\lim vec\left[E\left(S_{t,m}S'_{t,m}\right)\right]$, since $E\left(A_t^{\otimes 2}\right)$ and $E\left(vec\left(B_tB'_t\right)\right)$ are S-periodic, as

$$\begin{aligned} &\lim_{m\longrightarrow\infty} vec\left[E\left(S_{t,m}S'_{t,m}\right)\right] = \\ &\sum_{j=0}^{\infty}\left(\prod_{l=0}^{S-1} E\left(A_{n^{(t)}-l}^{\otimes 2}\right)\right)^{h^{(j)}}\left(\prod_{l=0}^{n^{(j)}-1} E\left(A_{n^{(t)}-l}^{\otimes 2}\right)\right) vecE\left(B_dB'_d\right), \end{aligned} \tag{4.3.10}$$

where $\tau^{(j)} = \left\lceil \frac{j-1}{S} \right\rceil$ and $d \in \{1, ..., S\}$ is such that $n^{(t)} - n^{(j)} = d + Sn$ with $n \in \mathbb{Z}$. Since $\prod_{l=0}^{n^{(j)}-1} E\left(A_{n^{(t)}-l}^{\otimes 2}\right)$ is uniformly bounded by $\max_{0\leq i\leq S}\left\|\prod_{l=0}^{S-1} E\left(A_{i-l}^{\otimes 2}\right)\right\|$ then a sufficient condition for (4.3.6) to hold is given by

$$\lim_{j\to\infty}\left\|\prod_{l=0}^{S-1} E\left(A_{n^{(t)}-l}^{\otimes 2}\right)\right\|^{[j/S]} = 0,$$

which accordingly holds if, and only if, the spectral radius of the matrix $\prod_{l=0}^{S-1} E\left(A_{n^{(t)}-l}^{\otimes 2}\right)$ (from where we obtain by a cyclic rearrangement the matrix $\prod_{s=0}^{S-1} E\left(A_{S-s}^{\otimes 2}\right)$) is less than one.

ii) To prove the necessity part, we have first to show that the periodic stationarity of model (4.3.2) implies that

$$\lim_{m\longrightarrow\infty}\sum_{j=0}^{m}\left(z'\otimes z'\right)\prod_{i=0}^{j-1}E\left(A_{t-i}^{\otimes 2}\right)vecE\left(B_{t-j}B_{t-j}'\right)<\infty \tag{4.3.11}$$

for all fixed $(p+q)\times 1$ vector z. Consider the notations (4.3.9) and let $V_t = E\left(Y_t Y_t'\right)$. Then since $E\left(S_{t,m}R_{t,m}'\right)=0$ we have

$$\begin{aligned}
vec\left(V_t\right) &= vec\left(E\left(S_{t,m}S_{t,m}'\right)\right)+vec\left(E\left(R_{t,m}R_{t,m}'\right)\right),\\
&= \sum_{j=0}^{m}\prod_{i=0}^{j-1}E\left(A_{t-i}^{\otimes 2}\right)vecE\left(B_{t-j}B_{t-j}'\right)+\prod_{i=0}^{m}E\left(A_{t-i}^{\otimes 2}\right)vecE\left(Y_{t-m-1}'Y_{t-m-1}\right),\\
&= \sum_{j=0}^{m}\prod_{i=0}^{j-1}E\left(A_{t-i}^{\otimes 2}\right)vecE\left(B_{t-j}B_{t-j}'\right)+\prod_{i=0}^{m}E\left(A_{t-i}^{\otimes 2}\right)vec\left(V_{t-m-1}\right).
\end{aligned}$$

Let

$$\left\{\begin{array}{ll} Q_{t-j,0}=E\left(B_{t-j}B_{t-j}'\right) \text{ and } Q_{t,j}=E\left(A_tQ_{t-1,j-1}A_t'\right) \\ R_{t-j,0}=V_{t-j} \qquad\qquad \text{ and } R_{t,j}=E\left(A_tR_{t-1,j-1}A_t'\right)\end{array}\right. \quad j=1,2,3...$$

Obviously the S-periodic matrices $Q_{t,j}$ and $R_{t,j}$, $j\in N$, are nonnegative definite for all $t\in\{1,...,S\}$ and we have

$$V_t=\sum_{j=0}^{m}Q_{t,j}+R_{t,m+1};\ \ m=0,1,2,...$$

If z is any $(p+q)\times 1$ fixed vector, then

$$\begin{aligned}
\infty &> z'V_tz=\sum_{j=0}^{m}z'Q_{t,j}z+z'R_{t,m+1}z,\\
&\geq \sum_{j=0}^{m}z'Q_{t,j}z,
\end{aligned}$$

for all $m\in N$ and $t\in\{1,...,S\}$. Thus $\sum_{j=0}^{m}z'^{\otimes 2}vecQ_{t,j}$ converge as $m\rightarrow\infty$ to a nonnegative matrix for z, which is exactly (4.3.11).

Now, we know that (4.3.11) implies that

$$\lim_{j\rightarrow\infty}W_{t,j}=\lim_{j\rightarrow\infty}\prod_{i=0}^{j-1}E\left(A_{t-i}^{\otimes 2}\right)vecE\left(B_{t-j}B_{t-j}'\right)=0, \text{ for all } t,$$

and hence

$$\lim_{j\to\infty}\left\|\prod_{i=0}^{j-1}E\left(A_{t-i}^{\otimes 2}\right)vecE\left(BB'\right)\right\|=0.$$

Therefore, all subsequences of sequence $\|W_{t,j}\|$ converge to 0, as $j\longrightarrow\infty$, in particular when j is a multiple of S, i.e.

$$\lim_{j\to\infty}\|W_{t,Sj}\|=0,$$

for all fixed t. Thus, adopting the notation (4.3.10) for the latter equation we get (4.3.8) as required. This completes the proof. ■

Note that condition (4.3.7) can be simplified by rewriting the matrix A_t as a block triangular matrix as follows

$$A_t=\begin{pmatrix}\Phi_t & \Lambda_t\\ \mathbf{0} & \Delta_t\end{pmatrix}$$

where

$$\begin{aligned}
\Phi_t &= \begin{pmatrix}\phi_{t,1} & \cdots & \phi_{t,p}\\ I_{(p-1)\times(p-1)} & & 0_{(p-1)\times 1}\end{pmatrix},\\
\Lambda_t &= \begin{pmatrix}\delta_{t,1}\cdots\delta_{t,q}\\ 0_{(p-1)\times q}\end{pmatrix},\\
\Delta_t &= \begin{pmatrix}\delta_{t,1} & \cdots & \delta_{t,q}\\ I_{(q-1)\times(q-1)} & & 0_{(q-1)\times 1}\end{pmatrix}.
\end{aligned}$$

Thus we have the following corollary.

Corollary 4.3.2 (Aknouche and Guerbyenne, 2005)

The $PAR-PARCH_S(p,q)$ model (4.3.2) *admits a unique $\mathcal{F}_t$ −measurable strictly periodically stationary solution of the form* (4.3.6) *if*

$$\rho\left(\prod_{s=0}^{S-1}\Phi_{S-s}\right)<1, \tag{4.3.12a}$$

and

$$\rho\left(\prod_{s=0}^{S-1}E\left(\Delta_{S-s}^{\otimes 2}\right)\right)<1. \tag{4.3.12b}$$

Proof

We have

$$\prod_{s=0}^{S-1} E\left(A_{S-s}^{\otimes 2}\right) = \prod_{s=0}^{S-1} E\left[\left(\begin{array}{cc} \Phi_{S-s} & \Lambda_{S-s} \\ \mathbf{0} & \Delta_{S-s} \end{array}\right) \otimes \left(\begin{array}{cc} \Phi_{S-s} & \Lambda_{S-s} \\ \mathbf{0} & \Delta_{S-s} \end{array}\right)\right]$$

$$= \left(\begin{array}{cc} \prod_{s=0}^{S-1} E\left(\Phi_{S-s} \otimes \left(\begin{array}{cc} \Phi_{S-s} & \Lambda_{S-s} \\ 0 & \Delta_{S-s} \end{array}\right)\right) & K \\ \mathbf{0} & \prod_{s=0}^{S-1} E\left(\Delta_{S-s+1} \otimes \left(\begin{array}{cc} \Phi_{S-s} & \Lambda_{S-s} \\ 0 & \Delta_{S-s} \end{array}\right)\right) \end{array}\right)$$

where K is a block matrix of suitable elements whose expressions are unimportant to us. Since

$$\prod_{s=0}^{S-1} E\left(\Phi_{S-s+1} \otimes \left(\begin{array}{cc} \Phi_{S-s} & \Lambda_{S-s} \\ 0 & \Delta_{S-s} \end{array}\right)\right) = \prod_{s=0}^{S-1} \Phi_{S-s} \otimes \left(\begin{array}{ccc} \Phi_{S-s} & \mathbf{0} & \\ \mathbf{0} & \begin{array}{ccc} 0 & \ldots & 0 \\ I_{(q-1)\times(q-1)} & & \mathbf{0} \end{array} & \Delta_{S-s} \end{array}\right),$$

then by a direct calculation, we obtain

$$\det\left[\lambda I_{(p+q)^2} - \prod_{s=0}^{S-1} E\left(\Phi_{S-s} \otimes \left(\begin{array}{cc} \Phi_{S-s} & \Lambda_{S-s} \\ 0 & \Delta_{S-s} \end{array}\right)\right)\right] = \lambda^{pq} \det\left[\lambda I_{p^2} - \prod_{s=0}^{S-1} \Phi_{S-s}^{\otimes 2}\right]$$

Similarly, we can show that

$$\det\left[\lambda I - \prod_{s=0}^{S-1} E\left(\Delta_{S-s} \otimes \left(\begin{array}{cc} \Phi_{S-s+1} & \Lambda_{S-s} \\ 0 & \Delta_{S-s} \end{array}\right)\right)\right] = \lambda^{pq} \det\left[\lambda I_{q^2} - \prod_{s=1}^{S} E\left(\Delta_{S-s}^{\otimes 2}\right)\right].$$

Therefore,

$$\rho\left(\prod_{s=0}^{S-1} E\left(A_{S-s}^{\otimes 2}\right)\right) < 1,$$

if, and only if,

$$\rho\left(\prod_{s=0}^{S-1} \Phi_{S-s}^{\otimes 2}\right) < 1,$$

and

$$\rho\left(\prod_{s=0}^{S-1} E\left(\Delta_{S-s}^{\otimes 2}\right)\right) < 1.$$

Note that

$$\rho\left(\prod_{s=0}^{S-1}\Phi_{S-s}^{\otimes 2}\right)=\rho\left(\prod_{s=0}^{S-1}\Phi_{S-s}\right)^{\otimes 2}<1,$$

(see for example Ling (1999a, Lemma 2.2)) if and only if

$$\rho\left(\prod_{s=0}^{S-1}\Phi_{S-s}\right)<1,$$

as required. ■

Remark 4.3.1

i) For the particular periodic $ARCH$ $(PARCH)$ model given by Bollerslev and Ghysels (1996), which is obtained from model (4.3.2) when $\phi_{tj}=0$, $j=1,...,p$ for all t, the periodic stationarity condition is given by (4.3.12b). This condition is the same as that given by Bollerslev and Ghysels (1996) for the particular $PARCH$ of period 2 and order 1. i.e. $\rho\left(\prod\limits_{s=1}^{S}\alpha_{s1}\right)=\rho\left(\prod\limits_{s=1}^{S}E\left(\delta_{s1}^{2}\right)\right)<1$.

i) The periodic AR model (PAR) which one obtains when $\delta_{t,i}=0$, $i=1,...,q$, is periodically stationary if and only if (4.3.12a) holds. This condition has been already given in Chapter 1 (cf, Proposition 1.2.2).

4.3.4 Existence of Higher-order moments

The existence of finite-order moments for a time series model is primordial for statistical inferences. Much attention has been given to this important point for the class of conditionally heteroskedastic models (cf., Engle, 1982; Milhoj, 1985; Chen and An, 1998; Ling, 1999; Ling and McAleer, 2002) because on the one hand, the implied conditions on the parameters are most "stringent" than the stability conditions and, on the other hand, the inference procedures require the existence of moments of a higher order for $ARCH$-type models than for traditional linear models. In what follows, a sufficient condition for the existence of higher-order moments of the process generated by (4.3.2) is given.

Theorem 4.3.2 (Aknouche and Guerbyenne, 2005)

Suppose that $E\left(e_t^{2r}\right)$ *is finite for all* $t \in Z$ *and* $r > 1$ *and that*

$$\prod_{s=0}^{S-1} E\left(A_{S-s}^{\otimes 2r}\right) < 1. \tag{4.3.13}$$

Then under the aforementioned assumptions, the processes generated by (4.3.2) *admits a unique periodically stationary* $(y_t,\, \varepsilon_t)$ *solution with* $E\left(y_t^{2r}\right) < \infty$ *and* $E\left(\varepsilon_t^{2r}\right) < \infty$ *for all* t.

Proof

According to the weak convergence theory (e.g. Billingsley, 1968) it is well known that a sufficient condition for the existence of $E\left(\varphi(X)\right)$ is that $\lim_{\tau\to\infty} \inf E\varphi(X_\tau)$ is finite, where (X_τ) is a sequence of random variables which converge in distribution to the random variable X and $\varphi(.)$ is a measurable function. This result is used here to prove this theorem. Define $Y_t^{(n)} = B_t + \sum_{j=1}^{n-1} \prod_{i=0}^{j-1} A_{t-i} B_{t-j}$ for $n > 0$ and $Y_t^{(0)} = 0$. It is clear that $Y_t^{(n)}$ verifies (4.3.2) and that $Y_t^{(n)}$ converges $a.s.$ to the unique solution of (4.3.2) given by (4.3.6) as $n \longrightarrow \infty$. Let

$$\pi_n(t) = Y_t^{(n)} - Y_t^{(n-1)} = \begin{cases} \prod_{i=0}^{n-1} A_{t-i} B_{t-n+1} = A_t \pi_{n-1}(t-1) & n \geq 1, \\ B_t & n = 0, \end{cases}$$

then using the properties of Kronecker product we get

$$\pi_n^{\otimes 2r}(t) = A_t^{\otimes 2r} \pi_{n-1}^{\otimes 2r}(t-1),$$

for some $r > 0$ and

$$\begin{aligned} E\left(\pi_n^{\otimes 2r}(t)\right) &= E\left(A_t^{\otimes 2r}\right) E\left(\pi_{n-1}^{\otimes 2r}(t-1)\right), \\ &= \prod_{i=0}^{n-1} E\left(A_{t-i}^{\otimes 2r}\right) E\left(B_{t-n+1}^{\otimes 2r}\right), \end{aligned} \tag{4.3.14}$$

which is a deterministic S-periodic difference equation of order 1. It is well known (e.g. Bentarzi and Hallin, 1994; Bentarzi, 1998; Bibi, 2004) that, since $E\left(B_{t-n+1}^{\otimes 2r}\right)$ is finite, a sufficient condition for $E\left(\pi_n^{\otimes 2r}(t)\right)$ to have a finite limit as $n \longrightarrow \infty$ is that (4.3.13) holds. Hence because $y_t^{(n)}$ converges to y_t where $y_t^{(n)}$ is the first entry of $Y_t^{(n)}$ and $\lim_{n\longrightarrow\infty} E\left(\left(y_t^{(n)} - y_t^{(n-1)}\right)^{\otimes 2r}\right)$ is finite, then using the weak convergence theory it follows that $E\left(y_t^{2r}\right) < \infty$; the same argument can be used for $E\left(\varepsilon_t^{2r}\right)$. This completes the proof.

4.4 Conclusion and future research

In this chapter we have attempted to give some probabilistic and algebraic properties of some conditionally heteroskedastic models in which coefficients are not more strictly stationary but are periodically stationary. In particular, we have explored, the causality and invertibility, the existence of a strictly stationary solution and the existence of finite moments for periodic $GARCH$ and double periodic AR models. A key role in our studies is played the Markovian formulation (with periodically distributed random coefficients) of the underlying models. The proposed formulations contain some useful models such as $ARCH$, Augmented $ARCH$, $AR-ARCH$, periodic AR and as special cases. As a possible future study, it would be of interest to introduce some flexible and parsimonious formulations that generalize the studied models. In particular we mention the important class of periodic $ARMA$ with periodic $GARCH$ models ($PARMA-GARCH$), the class of periodic RCA models, and the class of conditionally heteroskedastic $ARMA$ models ($PCHARMA$).

Chapter 5

On-line estimation of periodic $GARCH$ models

5.1 Introduction

The class of Autoregressive Conditional Heteroskedastic models ($ARCH$), introduced by Engle (1982) and their various extensions ($GARCH$, $FIGARCH$, $PGARCH$..., denoted here by $ARCH$-type) have proved their usefulness and appropriateness for modelling economic time series models, particularly the financial ones (stock exchange, rate of exchange,...) which are characterized by an instantaneous volatility feature that cannot be accounted for by the standard time-invariant time series models. Existing estimation methods for such models (see e.g. Weiss, 1986; Pantula, 1988; Lumsdaine, 1996; Bose and Mukherjee, 2001; Ling and McAleer, 2003; Berkes et al., 2003; Berkes and Horvath, 2004; Francq and Zakoian, 2004) are based on a beforehand fixed sample size thereby ignoring the sequential character of certain data, in particular financial data that are observed on a high-frequency basis and so, are progressively updated in real time. Furthermore, it was theoretically proved that the $ARCH$-type models are good approximations for certain continuous time series models (see Nelson 1990; Nelson and Foster, 1994; Drost and Werker, 1996). These facts are a motivation to search for estimation methods invoking these interesting features.

The main goal of this chapter is to develop some recursive estimation methods based on a two-stage least squares scheme ($2S-LS$) for estimating $PGARCH$ models. These methods, inspired from the automatic control field (cf, Chapter 3) provide consistent and asymptotically Gaussian estimators and generalize similar methods developped for $GARCH$ models (cf., Aknouche and Guerbyenne, 2006). Firstly, we generalize the *off-line two-stage least squares* ($2S-LS$) method proposed by Bose and Mukherjee (2003) for $ARCH$ models (see also Pantula, 1988 for the $ARCH(1)$ case) to their periodic counterparts. The obtained off-line $2S-LS$ method will be adapted to a recursive on-line context. We call the resulting algorithm the *two-stage RLS* algorithm ($2S-RLS$). Then, recursive estimation for $PGARCH$ models is considered. We propose two algorithms that parallel the ones available for standard linear models (cf, Chapter 3). The first one is the Two-Stage Pseudo-Linear Regression ($TS-PLR$) which is a version of the well-known Pseudo-Linear Regression (PLR) approach given in two stages; while the second one, namely the Two-Stage Recursive Predictor Error algorithm ($TS-RPE$) is an adaptation of the traditional Recursive Maximum Likelihood (RLM) method to deal with $PGARCH$ models. It will be shown that the obtained estimators are strongly consistent and, in particular, the RPE algorithm provides asymptotically Gaussian estimates.

The rest of this chapter is organized as follows. Section 2 extends the two-stage least squares ($2S$-LS) method proposed by Bose and Mukherjee (2003) and Pantula (1988) to the $PARCH$ case. In the third section, we develop from the latter off-line method, a two-stage recursive least squares algorithm ($2S$-RLS). This recursive method will be the basis of the other proposed algorithm for $PGARCH$ models. In section 4, we develop a $TS-PLR$ algorithm for $PGARCH$ model. Based on the Stochastic Lyapunov function approach (cf, Solo, 1979; Moore and Ledwich, 1980), we show the consistency of the obtained estimator. Section 5 develops a $TS-RPE$ algorithm. By means of the same approach, we show that the obtained estimator is asymptotically Gaussian. Finally, the performance of the proposed algorithms will be assessed by means of simulations.

5.2 Two-Stage Least Squares estimation of periodic $ARCH$ models

Consider the $PARCH$ process $\{\varepsilon_t;\ t\in\mathbb{Z}\}$ (cf, Chapter 4) given by

$$\begin{cases} \varepsilon_t = \sqrt{h_t}\zeta_t, \\ E(\varepsilon_t^2/\mathcal{F}_{t-1}) = h_t = \alpha_{t,0} + \sum\limits_{i=1}^{q} \alpha_{t,i}\,\varepsilon_{t-i}^2,\ t\in\mathbb{Z}, \end{cases} \tag{5.2.1}$$

where $\{\zeta_t, t\in\mathbb{Z}\}$ are independent and identically distributed ($i.i.d.$) of zero mean and unit variance and finite fourth moments.

Defining the innovation process $v_t = \varepsilon_t^2 - h_t = h_t\eta_t$, where $\eta_t = \zeta_t^2 - 1$, and letting $t = s + S\tau$, $s = 1, ..., S$, $\tau\in\mathbb{Z}$, one can rewrite equation (5.2.1) in the equivalent periodic autoregressive (PAR)

form for the process $\{\varepsilon_t^2, t \in \mathbb{Z}\}$ with respect to v_t

$$(1-\alpha_s(L))\ \varepsilon_{s+S\tau}^2 = \alpha_{s,0} + \ v_{s+S\tau},\ \ s = 1, ..., S \text{ and } \tau \in \mathbb{Z}, \tag{5.2.2}$$

where $\alpha_s(L) = \sum_{i=1}^{q} \alpha_{s,i}\ L^i$.

Bose and Mukherjee (2003) have developed a two-stage procedure to estimate the $ARCH$ parameters by solving two systems of linear equations. The method has been previously proposed by Pantula (1988) for the particular first order $ARCH(1)$. They have shown that the obtained estimators are asymptotically Gaussian and that the main estimator performs better than the quasi$-$maximum likelihood one, even for small samples. In this section, we develop a version of this method to deal with periodic $ARCH$ models. Consider the following regressor

$$\xi_{s+S\tau} = \left(\underbrace{\mathbf{0}', ..., \mathbf{0}'}_{s-1 \text{ time}}, \varepsilon_{s+S\tau-1}^2, \varepsilon_{s+S\tau-2}^2, ..., \varepsilon_{s+S\tau-q}^2, \underbrace{\mathbf{0}', ..., \mathbf{0}'}_{S-s \text{ time}}\right), \tag{5.2.3}$$

by means of which we rewrite model (5.2.2) in the linear regression form :

$$\varepsilon_{s+S\tau}^2 = \ \xi_{s+S\tau}'\alpha + h_{s+S\tau}(\alpha)\eta_{s+S\tau}, \tag{5.2.4}$$

where $\mathbf{0}$ is the $q \times 1$ null vector and where

$$\begin{array}{l} h_{s+S\tau}(\underline{\alpha}) = \underline{\xi}_{s+S\tau}'\underline{\alpha}, \\ \alpha' = (\alpha_1',\ \alpha_2',\ ...,\ \alpha_S')_{S(q+1)\times 1}, \\ \text{and} \\ \alpha_s = (\alpha_{s0}, \alpha_{s,1}, ..., \alpha_{s,q})'. \end{array}$$

Let $\varepsilon_{1-q}^2, ..., \varepsilon_0^2$ be the starting up values, which are considered here to be known, and $\varepsilon_1^2, \varepsilon_2^2, ..., \varepsilon_T^2$ the observed values of the process $\left\{\varepsilon_t^2, t \in \mathbb{Z}\right\}$, where $T = n + SN$, $n = 1, 2, ..., S$. To estimate α on the basis of the given values $\varepsilon_{1-q}^2, ..., \varepsilon_0^2, \varepsilon_1^2, \varepsilon_2^2, ..., \varepsilon_T^2$,, we consider in the first stage $h_{s+S\tau}(\alpha)$ as constant, while omitting its dependence on the unknown parameters α, which need to be estimated. In this situation, the ordinary least squares method provides the following pre-estimator

$$\widetilde{\alpha}(T) = \left(\sum_{k=1}^{T} \xi_k \xi_k'\right)^{-1} \sum_{k=1}^{T} \xi_k \varepsilon_k^2. \tag{5.2.5a}$$

(When the specification of the sample size is not paramount, $\widetilde{\alpha}(T)$ would be denoted simply by $\widetilde{\alpha}$). Obviously, $\widetilde{\alpha}(T)$ does not take into account the heteroskedasticity feature of the innovation components and hence it is less efficient than the quasi-maximum likelihood estimator. However, it can be improved (see Bose and Mukherjee, 2003). To achieve this improvement in its efficiency, we first rewrite model (5.2.4) in the form :

$$\frac{\varepsilon_{s+S\tau}^2}{h_{s+S\tau}(\alpha)} = \left(\frac{\xi_{s+S\tau}}{h_{s+S\tau}(\alpha)}\right)' \alpha + \eta_{s+S\tau},$$

and we replace the unknown vector α in $h_{s+S\tau}(\alpha)$ by its pre-estimation $\widetilde{\alpha}(T)$ obtained previously. Thus, we have the new "*quasi*" linear regression model

$$\frac{\varepsilon_{s+S\tau}^2}{h_{s+S\tau}(\widetilde{\alpha}(T))} = \left(\frac{\xi_{s+S\tau}}{h_{s+S\tau}(\widetilde{\alpha}(T))}\right)' \alpha + \eta_{s+S\tau}.$$

Applying again the ordinary least squares estimation method and being unaware of the presence of the stochastic component $\widetilde{\alpha}(T)$, we obtain a new estimate $\widehat{\alpha}(T)$ that we call the main estimator :

$$\widehat{\alpha}(T) = \left(\sum_{k=1}^{T} \frac{1}{h_k^2(\widetilde{\alpha}(T))} \xi_k \xi_k'\right)^{-1} \sum_{k=1}^{T} \frac{1}{h_k^2(\widetilde{\alpha}(T))} \xi_k \varepsilon_k^2. \tag{5.2.5b}$$

To study the probabilistic properties of the previous estimators, consider the expressions

$$F_1 = \sum_{s=1}^{S} E\left(\xi_s \xi_s'\right), \tag{5.2.6a}$$

$$F_2 = \sum_{s=1}^{S} E\left(\xi_s \xi_s' \left(\xi_s \alpha^*\right)^2\right), \tag{5.2.6b}$$

$$F_3 = \sum_{s=1}^{S} E\left(\xi_s \xi_s' \left(\xi_s \alpha^*\right)^{-2}\right), \tag{5.2.6c}$$

where α^* is the true value of the parameter α, and the assumptions :

$D1$: The process is strictly periodically stationary (cf, Chapter 4) and periodically ergodic.

$D2$: The moments $E\left(\varepsilon_i^2 \varepsilon_j^2 \varepsilon_k^2 \varepsilon_l^2\right)$ exist and are finite for all i, j, k, l.

$D3$: $E\left(\frac{\varepsilon_{-i}^2 \varepsilon_{-j}^2 \varepsilon_{-k}^2}{\left(\xi_s' \alpha^*\right)^3}\right) < \infty$ for each $1 \leq i, j, k \leq q$.

The same assumptions were considered by Bose and Mukherjee (2003) in the stationary case. Using the previous notations and assumptions, we can state the following theorem, which establishes, under the three previous assumptions, the asymptotic normality of $\widetilde{\alpha}$ and $\widehat{\alpha}$.

Theorem 5.2.1 (Bentarzi and Aknouche, 2004)

i) Under $D1$ and $D2$, we have

$$\sqrt{N}(\widetilde{\alpha}(T)-\alpha^*) \xrightarrow{D} N(0, Var(\eta_s^2)F_1^{-1}F_2F_1^{-1}). \tag{5.2.7}$$

ii) If, in addition, $D3$ holds then,

$$\sqrt{N}(\widehat{\alpha}(T)-\alpha^*) \xrightarrow{D} N(0, Var(\eta_s^2)F_3^{-1}). \tag{5.2.8}$$

Proof

Combining (5.2.5a) and (5.2.4) with α^* substituting α, and neglecting the sums

$$\frac{1}{N}\sum_{s=1}^{n}\xi_{s+SN}\xi'_{s+SN},$$

and

$$\frac{1}{\sqrt{N}}\sum_{s=1}^{n}h_{s+SN}(\alpha^*)\eta_{s+SN}\xi_{s+SN},$$

which are $o_p(1)$ as $N\to\infty$, we then obtain

$$\sqrt{N}(\widetilde{\alpha}\,(T)-\alpha^*) = \left(\sum_{s=1}^{S}\frac{1}{N}\sum_{m=0}^{N-1}\xi_{s+Sm}\xi'_{s+Sm}\right)^{-1}\sum_{s=1}^{S}\frac{1}{\sqrt{N}}\sum_{m=0}^{N-1}h_{s+Sm}(\alpha^*)\eta_{s+Sm}\xi_{s+Sm}.$$

From $D1$ and $D2$, we have

$$\frac{1}{N}\sum_{m=0}^{N-1}\xi_{s+Sm}\xi'_{s+Sm} \longrightarrow E\xi_s\xi'_s, \tag{5.2.9}$$

for all $s=1,...,S$. Moreover, we can easily verify that the applicability conditions of the Martingale central limit theorem (cf. Hall and Heyde 1980, Chapter 3 and Bose and Mukherjee, 2003) are satisfied. Therefore,

$$\frac{1}{\sqrt{N}}\sum_{m=0}^{N-1}h_{s+Sm}(\alpha^*)\eta_{s+mS}\xi_{s+Sm} \xrightarrow{D} N\left(0, Var(\eta_s^2)E\left[\xi_s\xi'_s h_s^2(\alpha^*)\right]\right), \tag{5.2.10}$$

for any $s=1,...,S$. Combining (5.2.9) and (5.2.10), we straightforwardly obtain (5.2.7).

To prove (5.2.8), we again replace $\varepsilon^2_{s+S\tau}$ in (5.2.5b) by its expression (5.2.2), for the true value α^*. Neglecting the "$o_p(1)$" sums

$$\frac{1}{N}\sum_{s=1}^{n}\frac{\xi_{s+SN}\xi'_{s+SN}}{h^2_{s+SN}(\widetilde{\alpha})},$$

and

$$\frac{1}{\sqrt{N}}\sum_{s=1}^{n}\frac{h_{s+SN}(\alpha^*)\eta_{s+SN}}{h^2_{s+SN}(\widetilde{\alpha})}\xi_{s+SN},$$

appearing in the obtained expression, we find

$$\sqrt{N}(\widehat{\alpha}\,(T)-\alpha^*)=\left(\sum_{s=1}^{S}\frac{1}{N}\sum_{m=0}^{N-1}\frac{\xi_{s+Sm}\xi'_{s+Sm}}{h^2_{s+Sm}(\widetilde{\alpha})}\right)^{-1}\sum_{s=1}^{S}\frac{1}{\sqrt{N}}\sum_{m=0}^{N-1}\frac{h_{s+Sm}(\alpha^*)\eta_{s+mS}}{h^2_{s+Sm}(\widetilde{\alpha})}\xi_{s+Sm}.$$

Applying the same techniques used in Bose and Mukherjee (2003), one can easily prove

$$\sum_{s=1}^{S}\frac{1}{N}\sum_{m=0}^{N-1}\left(\frac{1}{h^2_{s+Sm}(\widetilde{\alpha})}-\frac{1}{h^2_{s+Sm}(\alpha^*)}\right)\xi_{s+Sm}\xi'_{s+Sm}=o_p(1), \tag{5.2.11}$$

and

$$\sum_{s=1}^{S}\frac{1}{\sqrt{N}}\sum_{m=0}^{N-1}\left(\frac{1}{h^2_{s+Sm}(\widetilde{\alpha})}-\frac{1}{h^2_{s+Sm}(\alpha^*)}\right)h_{s+Sm}(\alpha^*)\eta_{s+mS}\xi_{s+Sm}=o_p(1). \tag{5.2.12}$$

Therefore, we obtain

$$\sqrt{N}(\widehat{\alpha}\,(T)-\alpha^*)=\left(\sum_{s=1}^{S}\frac{1}{N}\sum_{m=0}^{N-1}\frac{\xi_{s+Sm}\xi'_{s+Sm}}{h^2_{s+Sm}(\alpha^*)}\right)^{-1}\sum_{s=1}^{S}\frac{1}{\sqrt{N}}\sum_{m=0}^{N-1}\frac{\eta_{s+mS}}{h_{s+Sm}(\alpha^*)}\xi_{s+Sm}+o_p(1).$$

and applying again the Martingale central limit theorem (cf. Hall and Heyde, 1980), we deduce (5.2.8). ■

5.3 Two-Stage Recursive Least Squares algorithm

This section is concerned with the adaptation of the well-known recursive least squares (Plackett, 1950) algorithm (RLS) to cope with $PARCH$ models. The proposed method can, of course, be applied to the particular $ARCH$ case. The main idea is to solve the normal least squares equations (5.2.5) recursively via the RLS method, in two stages. In the first stage we obtain a preliminary recursive estimate $\widetilde{\alpha}_t$, which corresponds to a recursive least squares solution of (5.2.5a) when ignoring the conditional heteroskedasticity of the process given by (5.2.1). This preliminary estimate is then used in the second stage to obtain an improved recursive estimator $\widehat{\alpha}_t$, based on a modified form of (5.2.5b), which take into account the conditional heteroskedasticity by means of appropriate weights.

Clearly, the pre-estimator $\widetilde{\alpha}\,(T)$ given by (5.2.5a) has a standard OLS form and can be obtained recursively. Indeed, it is well known that (cf, Section 3.3) setting $\widetilde{R}_t = \sum\limits_{k=1}^{t} \xi_k \xi_k'$ and $\widetilde{\alpha}_t = \widetilde{\alpha}\,(t)$, the system (5.2.5$a$) can be solved recursively via the RLS algorithm, (e.g. Ljung and Söderström (1983, p. 17)), which provides the following recursive equations

$$\begin{aligned}
\widetilde{\alpha}_t &= \widetilde{\alpha}_{t-1} + \widetilde{R}_t^{-1} \xi_t \left(\varepsilon_t^2 - \xi_t' \widetilde{\alpha}_{t-1} \right), && (5.3.1a) \\
\widetilde{R}_t &= \widetilde{R}_{t-1} + \xi_t \xi_t', \qquad t \geq 2, && (5.3.1b)
\end{aligned}$$

with starting values $\widetilde{R}_1 = \xi_1 \xi_1'$ and $\widetilde{\alpha}_1 = \widetilde{R}_1^{-1} \xi_1 \varepsilon_1^2$ provided that $\widetilde{R}_1$ is nonsingular.

Unfortunately, (5.2.5b) cannot be solved recursively as (5.3.5a) because of the presence of $\widetilde{\alpha}(T)$ in the weights $h_k^{-2}(\widetilde{\alpha}(T))$, $k = 1, ..., NS$. However, it is possible to slightly modify (5.2.5b) to obtain an adequate form for a recursive resolving method, while preserving the same asymptotic properties of the main estimator $\widehat{\alpha}\,(T)$. So, we propose the following modified equation

$$\left(\sum_{k=1}^{t} \frac{1}{h_k^2(\widetilde{\alpha}_k)} \xi_k \xi_k' \right) \widehat{\alpha}_t = \sum_{k=1}^{t} \frac{\varepsilon_k^2}{h_k^2(\widetilde{\alpha}_k)} \xi_k, \tag{5.3.2}$$

in which the weight $h_k^{-2}(\widetilde{\alpha}(T))$ is replaced by $h_k^{-2}(\widetilde{\alpha}_k)$, where $\widehat{\alpha}_t$ is our main recursive estimator of α at time t.

It is obvious that (5.2.5b) and (5.3.2) give different estimators; however, they have the same asymptotic properties, as we will show below.

To solve (5.3.2) recursively we set $\widehat{R}_t = \sum\limits_{k=1}^{t} \dfrac{1}{h_k^2(\widetilde{\alpha}_{k-1})} \xi_k \xi_k'$. Then, we have the following recursive equations

$$\widehat{\alpha}_t = \widehat{\alpha}_{t-1} + \frac{\widehat{R}_t^{-1}\xi_t\left(\varepsilon_t^2 - \xi_t'\widehat{\alpha}_{t-1}\right)}{\left(\xi_t'\widetilde{\alpha}_t\right)^2}, \tag{5.3.3a}$$

$$\widehat{R}_t = \widehat{R}_{t-1} + \frac{\xi_t\xi_t'}{\left(\xi_t'\widetilde{\alpha}_t\right)^2}, \qquad t \geq 2, \tag{5.3.3b}$$

with starting values $\widehat{R}_1 = \frac{1}{h_1^2(\widetilde{\alpha}_1)}\xi_1\xi_1'$ and $\widehat{\alpha}_1 = \widehat{R}_1^{-1}\xi_1\frac{\varepsilon_1^2}{h_1^2(\widetilde{\alpha}_1)}$ provided that $\widehat{R}_1$ is nonsingular. To avoid the singularity problem of $\widetilde{R}_1$ and $\widehat{R}_1$, the starting-up values $\widetilde{\alpha}_0$, $\widehat{\alpha}_0$, $\widetilde{R}_0$ and $\widehat{R}_0$ are fixed so that the recursive estimators will as close as possible to the off-line solutions (5.2.5). So, starting from the recursive equations (5.3.1) and (5.3.3) we shall (approximately) find (5.2.5). Multiplying both sides of (5.3.1a) by $\widetilde{R}_t$, using (5.3.1b) and summing the resulting equations side by side, for $k = 1$ to t we obtain :

$$\left(\widetilde{R}_0 + \sum_{k=1}^{t}\xi_k\xi_k'\right)\widetilde{\alpha}_{s+S\tau} = \widetilde{R}_0\widetilde{\alpha}_0 + \sum_{k=1}^{t}\varepsilon_k^2\xi_k. \tag{5.3.4a}$$

Similar manipulations applied to (5.3.3a) and (4.3.3b) give

$$\left(\widehat{R}_0 + \sum_{k=1}^{s+S\tau}\frac{\xi_k\xi_k'}{(\xi_k'\widetilde{\alpha}_k)^2}\right)\widehat{\alpha}_{s+S\tau} = \widehat{R}_0\widehat{\alpha}_0 + \sum_{k=1}^{s+S\tau}\frac{\varepsilon_k^2}{(\xi_k'\widetilde{\alpha}_k)^2}\xi_k. \tag{5.3.4b}$$

Comparing (5.2.5) with (5.3.4), a reasonable choice of the starting values is to put

$$\widetilde{\alpha}_0 = \widehat{\alpha}_0 = 0,$$

and

$$\widetilde{R}_0 = \widehat{R}_0 = M^{-1}I,$$

where M is a large enough positive real.

In order to improve the algorithmic complexity of the latter procedure, setting $\widetilde{P}_{s+S\tau} = \widetilde{R}_{s+S\tau}^{-1}$ and $\widehat{P}_{s+S\tau} = \widehat{R}_{s+S\tau}^{-1}$, and rewriting the set of equations (5.3.1) and (5.3.3) by using the well-known matrix inversion lemma (cf, Appendix A), we obtain the following algorithm which we call the two-stage RLS (2S-RLS).

Algorithm 5.3.1 The $2S - RLS$ algorithm

The two-stage recursive least squares algorithm ($2S$-RLS)*, for estimating* $PARCH$ *models, is given by the following set of recursive equations :*

$$\underbrace{\begin{cases} \widetilde{v}_t = \varepsilon_t^2 - \xi_t'\widetilde{\alpha}_{t-1}, \\ \widetilde{\alpha}_t = \widetilde{\alpha}_{t-1} + \dfrac{\widetilde{P}_{t-1}\xi_t\widetilde{v}_t}{1+\xi_t'\widetilde{P}_{t-1}\xi_t},\ \widetilde{\alpha}_0 = 0, \\ \widetilde{P}_t = \widetilde{P}_{t-1} - \dfrac{\widetilde{P}_{t-1}\xi_t\xi_t'\widetilde{P}_{t-1}}{1+\xi_t'\widetilde{P}_{t-1}\xi_t},\ \widetilde{P}_0 = M.I, \end{cases}}_{\text{first stage}} ; \underbrace{\begin{cases} \widehat{v}_t = \varepsilon_t^2 - \xi_t'\widehat{\alpha}_{t-1}, \\ \widehat{\alpha}_t = \widehat{\alpha}_{t-1} + \dfrac{\widehat{P}_{t-1}\xi_t\widetilde{v}_t}{\left(\xi_t'\widetilde{\alpha}_t\right)^2 + \xi_t'\widehat{P}_{t-1}\xi_t},\ \widehat{\alpha}_0 = 0, \\ \widehat{P}_t = \widehat{P}_{t-1} - \dfrac{\widehat{P}_{t-1}\xi_t\xi_t'\widehat{P}_{t-1}}{\left(\xi_t'\widetilde{\alpha}_t\right)^2 + \xi_t'\widehat{P}_{t-1}\xi_t},\ \widehat{P}_0 = M\,I. \end{cases}}_{\text{second stage}} \tag{5.3.5}$$

To study the asymptotic properties of the recursive estimators given by (5.3.5), we need to consider the following additional assumption :

$D4$: The expectation $E\left(\varepsilon_t^6\right)$ is finite for any t.

Note that a more stringent assumption that the eight-order moment of the $ARCH$ process exists, has been considered in Bose and Mukherjee (2003).

The following theorem shows that our $2S$-RLS algorithm provides estimators that have the same asymptotic distribution as the off-line ones proposed (cf, Section 5.2) and hence the main estimator is asymptotically efficient under the conditional normality assumption.

Theorem 5.3.1

i) Under assumptions $D1$ *the recursive estimators* $\widetilde{\alpha}_t$ *and* $\widehat{\alpha}_t$ *converge almost surely to the true value* α^* *of the parameter* α.

ii) In addition, if $D4$ *holds, we have :*

$$\sqrt{s+S\tau}(\widetilde{\alpha}_{s+S\tau} - \alpha^*) \underset{\tau\to\infty}{\overset{D}{\longrightarrow}} N(0, SVar(\eta_0^2)F_1^{-1}F_2F_1^{-1}), \tag{5.3.6}$$

and

$$\sqrt{s+S\tau}(\widehat{\alpha}_{s+S\tau} - \alpha^*) \underset{\tau\to\infty}{\overset{D}{\longrightarrow}} N\left(0, SVar(\eta_0^2)F_3^{-1}\right), \tag{5.3.7}$$

for all $s = 1, ..., S$, *where* F_1, F_2 *and* F_3 *are given by* (5.2.6).

Proof

Obviously, as $t \to \infty$, equations (5.3.4a) and (5.2.5a) are equivalent in the sense that the associated estimators have the same asymptotic properties. Hence, result (5.3.6) follows immediately from (5.2.7) (Theorem 5.2.1). By the same argument, equation (5.3.4b) is asymptotically equivalent to (5.3.2). Now, we analyze the asymptotic properties of the estimator obtained from this last equation. From (5.3.2) and (5.2.2) we have

$$t^{1/2}(\widehat{\alpha}_t - \alpha^*) = \left[t^{-1} \sum_{k=1}^{t} \frac{\xi_k \xi_k'}{(\xi_k' \widetilde{\alpha}_k)^2} \right]^{-1} \left[t^{-1/2} \sum_{k=1}^{t} \frac{(\xi_k' \alpha^*)\, \eta_k}{(\xi_k' \widetilde{\alpha}_k)^2} \xi_k \right]. \tag{5.3.8}$$

As in Bose and Mukherjee (2003), to establish (5.3.8), it is enough to show that

$$t^{-1} \sum_{k=1}^{t} \left(\frac{1}{(\xi_k' \widetilde{\alpha}_k)^2} - \frac{1}{(\xi_k' \alpha^*)^2} \right) \xi_k \xi_k' = o_p(1), \tag{5.3.9a}$$

and

$$t^{-1/2} \sum_{k=1}^{t} \left(\frac{1}{(\xi_k' \widetilde{\alpha}_k)^2} - \frac{1}{(\xi_k' \alpha^*)^2} \right) (\xi_k' \alpha^*)\, \eta_k \xi_k = o_p(1), \tag{5.3.9b}$$

and by the Martingale central limit theorem (cf, Hall and Heyde, 1980) we conclude that (5.3.7) is true. To prove (5.3.9a), we use the mean value theorem as in Bose and Mukherjee (2003). Then, we have

$$\begin{aligned} t^{-1} \sum_{k=1}^{t} \left(\frac{1}{(\xi_k' \widetilde{\alpha}_k)^2} - \frac{1}{(\xi_k' \alpha^*)^2} \right) \xi_k \xi_k' \quad &= \quad -2t^{-1} \sum_{k=1}^{t} \frac{(\widetilde{\alpha}_k - \alpha^*)' \xi_k}{\chi_k^3} \xi_k \xi_k' \\ &\overset{def}{=} -2S_t \end{aligned} \tag{5.3.10}$$

where the constant χ_k^3 is such that

$$\begin{aligned} 0 \quad &< \quad \frac{1}{\chi_k^3} \leq \frac{1}{\left(\xi_k' \alpha^*\right)^3} \left[1 + \frac{\xi_k' \alpha^*}{\xi_k' \widetilde{\alpha}_k} \right]^3 \\ &\leq \quad \frac{1}{\left(\xi_k' \alpha^*\right)^3} \left[1 + \left\{ 1 + \sum_{j=0}^{q} \frac{\alpha_{r,j}^*}{\widetilde{\alpha}_{r,j}^{(k)}} \right\} \right]^3 \leq (\alpha_0^*)^{-3} \left[2 + \sum_{j=0}^{q} \frac{\alpha_{r,j}^*}{\widetilde{\alpha}_{r,j}^{(k)}} \right]^3, \end{aligned} \tag{5.3.11}$$

in which the upper bound of the last inequality is bounded $a.s.$, for all k, in view of (5.3.6). Let $C_1 > 0$ be such a bounding constant. Note that in the first inequality of (5.3.11), we have used the relation (17) in Bose and Mukherjee (2003, p. 133), where the positive integer r is the remainder of the division of k by S.

Then from (5.3.10) and (5.3.11), we have

$$\|S_t\| \leq C_1 t^{-1} \sum_{k=1}^{t} \|\widetilde{\alpha}_k - \alpha^*\| \|\xi_k\|^3 .$$

Since the terms in the latter inequality satisfy

$$k^{-1} \|\widetilde{\alpha}_k - \alpha^*\| \|\xi_k\|^3 \leq C_2 k^{-3/2} \|\xi_k\|^3 ,$$

where from (5.3.6), C_2 is such that

$$k^{1/2} \|\widetilde{\alpha}_k - \alpha^*\| \leq C_2, \tag{5.3.12}$$

by assumption $D4$ we have

$$\sum_{k=1}^{\infty} k^{-3/2} E\left(\|\xi_k\|^3\right) < \infty,$$

and then (see Lukacs, 1975, p. 80)

$$\sum_{k=1}^{\infty} k^{-1} \|\widetilde{\alpha}_k - \alpha^*\| \|\xi_k\|^3 < \infty.$$

Thus, from Kronecker lemma (cf. Appendix B) (5.3.9a) follows.

To establish (5.3.9b) we again use the mean value theorem

$$\begin{aligned} t^{-1/2} \sum_{k=1}^{t} \left(\frac{1}{(\xi_k' \widetilde{\alpha}_k)^2} - \frac{1}{(\xi_k' \alpha^*)^2} \right) \left(\xi_k' \alpha^*\right) \eta_k \xi_k &= -2t^{-1/2} \sum_{k=1}^{t} \frac{\left(\left(\widetilde{\alpha}_k - \alpha^*\right)' \xi_k\right) \left(\xi_k' \alpha^*\right) \eta_k}{\chi_k^3} \xi_k, \\ &\stackrel{def}{=} -2t^{-\delta} S_{t,\delta}. \end{aligned}$$

To establish (5.3.9b) it suffices to show that

$$\begin{aligned} S_{t,\delta} &\stackrel{def}{=} t^{-1/2+\delta} \sum_{k=1}^{t} \frac{\left(\left(\widetilde{\alpha}_k - \alpha^*\right)' \xi_k\right) \left(\xi_k' \alpha^*\right) \eta_k}{\chi_k^3} \xi_k, \\ &= o_p(1). \end{aligned}$$

However, if we can show that

$$E\left(\|S_{t,\delta}\|\right) < \infty,$$

then by Markov inequality we have

$$\begin{aligned} P\left(t^{-\delta}\left\|S_{t,\delta}\right\| \geq \eta\right) &= P\left(\left\|S_{t,\delta}\right\| \geq t^{\delta}\eta\right), \\ &\leq \frac{E\left(\left\|S_{t,\delta}\right\|\right)}{t^{\delta}\eta}, \\ &= o_p(1), \end{aligned}$$

which is exactly (5.3.9b). But using (5.3.12) and assumption $D4$ we have

$$\begin{aligned} E\left(\left\|S_{t,\delta}\right\|\right) &\leq C_2 t^{-1/2+\delta} \sum_{k=1}^{t} k^{-1} \frac{1}{\chi_k^3} E\left(\left\|\xi_k\right\|^3\right) E\left|\eta_k\right|, \\ &\leq \frac{1}{\chi_k^3} C_2 t^{-1/2+\delta} \log t, \\ &< \infty, \end{aligned}$$

which completes the proof. ■

5.4 A Two-Stage Pseudo-Linear regression algorithm

In this section, we extend the pseudo linear regression method (cf, Chapter 3) to the $PGARCH$ case. We recall that this approach consists of writing a $PGARCH$ model (4.2.1) in a linear regression form with respect to the square of the process $\{\varepsilon_t^2, t \in \mathbb{N}\}$, where some regressors are unobservable; and then estimating, by the ordinary least squares method, both these unobservable components and the unknown parameters of the underlying model. However, this method does not provide efficient estimators, even under the conditional normality assumption. This is because the innovations, function of the conditional variance, are correlated with the past of the dependent variable ε_t^2. For this reason, we invoke the well-known two-stage least squares method which was introduced initially to estimate autoregressive models with $ARCH$ errors, by Pantula (1988) and recently used by Bose and Mukherjee (2003) for estimating $ARCH$ models.

Consider the $PGARCH_S(p,q)$ process $\{\varepsilon_t;\ t \in \mathbb{Z}\}$ given by

$$\begin{cases} \varepsilon_t = \sqrt{h_t}\zeta_t, \\ (1-\beta_t(L))h_t = \alpha_{t,0} + \alpha_t(L)\,\varepsilon_t^2,\ t \in \mathbb{Z}, \end{cases} \tag{5.4.1}$$

where $\{\zeta_t, t \in \mathbb{Z}\}$ is an independent and identically distributed ($i.i.d.$) sequence with zero mean and unit variance. The polynomials $\alpha_t(L)$ and $\beta_t(L)$ are given by : $\alpha_t(L) = \sum_{i=1}^{q} \alpha_{t,i} L^i$, $\beta(L) = \sum_{j=1}^{p} \beta_{t,j} L^j$, where L is the backward shift operator, and $\alpha_{t,i}$ and $\beta_{t,j}$ are periodic in time such that $\alpha_{t,i} \geq 0$, $\beta_{t,j} \geq 0$, $i = 0, ..., q$, $j = 1, ..., p$.

Letting $v_t = \varepsilon_t^2 - h_t = h_t \eta_t$, where $\eta_t = \zeta_t^2 - 1$, model (5.4.1) can be rewritten in a regression form, which will be used in the subsequent sections

$$\varepsilon_t^2 - \alpha_0 - \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 = v_t - \sum_{j=1}^{p} \beta_j \left(\varepsilon_{t-j}^2 - v_{t-j}(\theta) \right).$$

For the rest of this paper we suppose that model (5.4.1) is causal and invertible (cf. Chapter 4). We also suppose that the process $\{\varepsilon_t;\ t \in \mathbb{Z}\}$ is ergodic, that is $\{\varepsilon_{s+S\tau}\}_\tau$ is an ergodic process. It is possible to say under what conditions the $PGARCH$ process is periodically ergodic. This could be acheived by making a Volterra series expansion of ε_t^2 in terms of $\{\blacksquare_t^2\}_t$ (see for example Giraitis et al. (2000), where a Volterra expansion and moment conditions for $ARCH(\infty)$ processes are given) and using Theorem 3.5.8 of Stout (1974).

Letting θ, the prediction of ε_t^2 based on the available information up to time $t-1$ is given by

$$\widehat{\varepsilon}_t(\theta) = \varepsilon_t^2 - v_t(\theta),\ t \in \mathbb{Z}, \tag{5.4.2a}$$

where $v_t(\theta)$ is the prediction error function which can be recursively evaluated in terms of ε_{t-i}^2, $i = 1, ..., q$ and $v_{t-j}(\theta)$, $j = 1, ..., p$, from the following stochastic difference equation

$$v_t(\theta) = \varepsilon_t^2 - \alpha_{t,0} - \sum_{i=1}^{q} \alpha_{t,i} \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_{t,j} \left(\varepsilon_{t-j}^2 - v_{t-j}(\theta) \right),\ t \in \mathbb{Z}. \tag{5.4.2b}$$

Since the effects of initial values of (5.4.2b) are not much persisting for a causal model, one can without affecting the quality of prediction use zeros or other initial values (see e.g. Francq and Zakoïan, 2004). Combining (5.4.2a) and (5.4.2b), $\widehat{\varepsilon}_t(\theta)$ can be written in the recursive form

$$\widehat{\varepsilon}_t(\theta) = \sum_{j=1}^{p} \beta_j \widehat{\varepsilon}_{t-j}(\theta) + \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2,\ t \geq 1, \tag{5.4.3}$$

with fixed starting values.

Given a time series $\varepsilon^t = \{\varepsilon_1, ..., \varepsilon_t\}$ with a number of observations t which is not beforehand fixed, we consider the weighted least squares prediction error criterion

$$V_t(\theta) = \frac{1}{2}\sum_{k=1}^{t} \frac{v_k^2(\theta)}{h_k^2(\theta)}.$$

We mention that the use of the weights $\dfrac{1}{h_k^2(\theta)}$ in the latter squares error criterion is quite adequate because, as is mentioned above, the square of the innovation $v_t^2(\theta^*) = v_t^2 = h_t^2(\zeta^2 - 1)^2$, where θ^* is the true value of the unknown parameter θ, is conditionally correlated with the past of $\{\varepsilon_t^2\}$. The two recursive estimation algorithms which we present below, and which are obtained through the minimization of this expression have the smallest temporal and spatial algorithmic complexities as possible.

For $t = s + S\tau$, $s = 1, 2, ..., S$ and $\tau \in \mathbb{Z}$, let $\varphi_{s+S\tau}$ be the $S(p+q+1)\times 1$ "pseudo regressor" vectors defined as follows

$$\varphi_{s+S\tau} = \left(\underbrace{\mathbf{0}', ..., \mathbf{0}'}_{s-1 \text{ time}}, \underline{\varepsilon}'_{s,\tau}, -\underline{w}_{s,\tau}, \underbrace{\mathbf{0}', ..., \mathbf{0}'}_{S-s \text{ time}}\right), \tag{5.4.4}$$

where $\underline{0}$, $\underline{\varepsilon}_{s,\tau}$ and $\underline{w}_{s,\tau}$ are, respectively, the $(p+q+1)\times 1$ null vectors, the $q \times 1$ vectors of the squared process $\underline{\varepsilon}'_{s,\tau} = (\varepsilon^2_{s+S\tau-1}, \varepsilon^2_{s+S\tau-2}, ..., \varepsilon^2_{s+S\tau-q})'$, and the $p \times 1$ vectors of the unobservable components $\underline{w}'_{i,\tau} = (\varepsilon^2_{s+S\tau-1} - v_{s+S\tau-1}, ..., \varepsilon^2_{s+S\tau-p} - v_{s+S\tau-p})'$, $s = 1, 2, ..., S$ and $\tau \in \mathbb{Z}$. Using these notations, model (5.4.2) can be rewritten in the pseudo linear regression form :

$$\varepsilon_t^2 = \varphi_t'\theta + v_t. \tag{5.4.5}$$

The inconvenience of the model (5.4.5) is that the regressor vector φ_t contains unobserved components (innovations). Therefore, the unknown parameter vector cannot be forwardly estimated by the ordinary least squares method. However, if we suppose temporarily that these unobservable values are known, model (5.4.5) can then be seen as a linear regression model with periodically conditionally heteroskedastic errors, for which one can use the well-known recursive least squares algorithm (in two stages, namely $2S$-RLS) to obtain firstly, the preliminary estimate $\widetilde{\theta}_t$, which corresponds to the least squares solution, in a recursive form, when ignoring the heteroskedasticity of the errors. This preliminary estimate is then used to obtain an improved estimator $\widehat{\theta}_t$ that takes into account the conditional heteroskedasticity, by means of appropriate weights. Along the lines of Chapter 3 (Section 3.4) we obtain the following recursive equations

$$\underbrace{\begin{cases} \widetilde{v}_t = \varepsilon_t^2 - \widetilde{\theta}'_{t-1}\varphi_t, \\ \widetilde{\theta}_t = \widetilde{\theta}_{t-1} + \widetilde{R}_t^{-1}\varphi_t\widetilde{v}_t, \\ \widetilde{R}_t = \widetilde{R}_{t-1} + \varphi_t\varphi'_t, \end{cases}}_{\text{first stage}}, \underbrace{\begin{cases} \widehat{v}_t = \varepsilon_t^2 - \widehat{\theta}'_{t-1}\varphi_t, \\ \widehat{\theta}_t = \widehat{\theta}_{t-1} + \widehat{R}_t^{-1}\varphi_t \dfrac{\widehat{v}_t}{\left(\varphi'_t\widetilde{\theta}_t\right)^2}, \\ \widehat{R}_t = \widehat{R}_{t-1} + \dfrac{\varphi_t\varphi'_t}{\left(\varphi'_t\widetilde{\theta}_t\right)^2}, \end{cases}}_{\text{second stage}} \tag{5.4.6}$$

where $\widetilde{R}_t$ is an approximation of the design matrix of the regression problem (5.4.5) with respect to the time series $\{\varepsilon_k^2, \, k = 1, ..., t\}$ and $\widehat{R}_t$ is also the design matrix when dividing both sides of (5.4.5) by the weights $\dfrac{1}{h_k(\widetilde{\theta}_k)}$. Note that it is possible to improve the algorithmic complexity of the last algorithm by using the well-known matrix inversion lemma (MIL) (cf, Chapter 3).

Hitherto, the innovations (v_t) were supposed known. However, as they are unknown in practice, we can proceed to estimate them using an approximation of the expression (5.4.2) where the unknown parameter θ is replaced once by its current preliminary estimate $\widetilde{\theta}_t$ and for a second time, by the improved estimate $\widehat{\theta}_t$ obtained by the recursive algorithm (5.4.6). This gives

$$\widetilde{\vartheta}_t = \varepsilon_t^2 - \widetilde{\varphi}'_t\widetilde{\theta}_t, \tag{5.4.7a}$$

$$\widehat{\vartheta}_t = \varepsilon_t^2 - \widehat{\varphi}'_t\widehat{\theta}_t, \tag{5.4.7b}$$

where $\widetilde{\varphi}_t$ and $\widehat{\varphi}_t$ are obtained from (5.4.6) when replacing v_t by $\widetilde{\vartheta}_t$ and $\widehat{\vartheta}_t$ respectively. Using these notations, and combining the estimation of the unobservable components given by (5.4.7a) with that of the unknown parameters, one can propose a pseudo linear regression algorithm, in two stages, as follows.

Algorithm 5.4.1 ($TS - PLR$ algorithm)

The two-stage pseudo−linear regression ($TS-PLR$) estimation of a $PGARCH$ model is given by the following recursive equations

$$\underbrace{\left\{\begin{array}{l} \widetilde{v}_t = \varepsilon_t^2 - \widetilde{\varphi}_t'\widetilde{\theta}_{t-1}, \\ \widetilde{\theta}_t = \widetilde{\theta}_{t-1} + \dfrac{\widetilde{P}_{t-1}\widetilde{\varphi}_t\widetilde{v}_t}{1+\widetilde{\varphi}_t'\widetilde{P}_{t-1}\widetilde{\varphi}_t},\ \widetilde{\theta}_{-1} = 0, \\ \widetilde{P}_t = \widetilde{P}_{t-1} - \dfrac{\widetilde{P}_{t-1}\widetilde{\varphi}_t\widetilde{\varphi}_t'\widetilde{P}_{t-1}}{1+\widetilde{\varphi}_t'\widetilde{P}_{t-1}\widetilde{\varphi}_t},\ \widetilde{P}_{-1} = M.I, \end{array}\right.}_{\text{first stage}}, \underbrace{\left\{\begin{array}{l} \widehat{v}_t = \varepsilon_t^2 - \widehat{\varphi}_t'\widehat{\theta}_{t-1} \\ \widehat{\theta}_t = \widehat{\theta}_{t-1} + \dfrac{\widehat{P}_{t-1}\widehat{\varphi}_t\widehat{v}_t}{\left(\widetilde{\varphi}_t'\widetilde{\theta}_{t-1}\right)^2 + \widehat{\varphi}_t'\widehat{P}_{t-1}\widehat{\varphi}_t},\ \widehat{\theta}_0 = 0 \\ \widehat{P}_t = \widehat{P}_{t-1} - \dfrac{\widehat{P}_{t-1}\widehat{\varphi}_t\widehat{\varphi}_t'\widehat{P}_{t-1}}{\left(\widetilde{\varphi}_t'\widetilde{\theta}_{t-1}\right)^2 + \widehat{\varphi}_t'\widehat{P}_{t-1}\widehat{\varphi}_t},\ \widehat{P}_0 = MI \end{array}\right.}_{\text{second stage}} \tag{5.4.8}$$

where $\widetilde{P}_t = \widetilde{R}_t^{-1}$, $\widehat{P}_t = \widehat{R}_t^{-1}$ and M is a large enough number.

To study the convergence property of the estimators obtained by the algorithm (5.4.8), one can use one of the well-known approaches in the on-line estimation literature, such as the Ordinary Differential Equation approach (Ljung 1977*a*; 1977*b*), Hannan (1980)'s approach or the stochastic Lyapunov function approach due to Moore and Ledwich (1980) and Solo (1979). However, we adopt the last approach which is proved to assure the strong convergence property of the recursive estimator without being monitored. Consider the following assumptions :

$E1$: The periodically causal, invertible and ergodic $PGARCH$ process $\{\varepsilon_t, t \in \mathbb{Z}\}$ given by (5.4.2) has finite fourth moments, then satisfies

$$\lim_{t\to\infty} \sup \frac{1}{t}\sum_{k=1}^{t} \varepsilon_k^4 < \infty, \tag{5.4.9}$$

and ensures that $\lim\limits_{t\to\infty} \frac{1}{t}\sum\limits_{k=1}^{t} \widetilde{\varphi}_k\widetilde{\varphi}_k' = \widetilde{R}$ and $\lim\limits_{t\to\infty} \frac{1}{t}\sum\limits_{k=1}^{t} (\widetilde{\varphi}_k'\underline{\widetilde{\theta}}_k)^{-2}\widehat{\varphi}_k\widehat{\varphi}_k' = \widehat{R}$ are positive definite matrices.

$E2$: The filters $1-\beta_s(L)$ are such that

$$\mathrm{Re}\left[(1-\beta_s(e^{i\omega}))^{-1} - \frac{1}{2}\right] > 0 \quad \forall\omega \in \,]-\pi,\pi[\,,\text{ for all } s = 1,...,S. \tag{5.4.10}$$

Note that for the $PARMA$ case, Adams and Goodwin (1995) have replaced the condition (5.4.10) by the very strict passivity property (cf, Chapter 3, Hill and Moylan, 1980 and Moore and Ledwich, 1980) of a certain model. They have also given a sufficient condition for this model to be very strictly passive. Under the above assumptions, we can prove the following result.

Theorem 5.4.1

The TS-PLR obtained estimators $\widetilde{\theta}_t$ and $\widehat{\theta}_t$ converge almost surely to θ^*, in the sense that

$$\frac{1}{t}\left(\widetilde{\theta}_t-\theta^*\right)'\widetilde{R}_t\left(\widetilde{\theta}_t-\theta^*\right)\xrightarrow{a.s.}0,\ as\ t\to\infty, \tag{5.4.11a}$$

$$\frac{1}{t}\sum_{k=1}^{t}\left(\widetilde{\vartheta}_k-\upsilon_k\right)^2\xrightarrow{a.s.}0,\ as\ t\to\infty, \tag{5.4.11b}$$

and

$$\frac{1}{t}\left(\widehat{\theta}_t-\theta^*\right)'\widehat{R}_t\left(\widehat{\theta}_t-\theta^*\right)\xrightarrow{a.s.}0,\ as\ t\to\infty, \tag{5.4.12a}$$

$$\frac{1}{t}\sum_{k=1}^{t}\left(\widehat{\vartheta}_k-\upsilon_k\right)^2\xrightarrow{a.s.}0,\ as\ t\to\infty. \tag{5.4.12b}$$

Proof

The proof of this theorem is based on the stochastic Lyapunov function approach as in Chapter 3 (see also Solo, 1979 and Moore and Ledwich, 1980). Results (5.4.11) can be easily shown by similar techniques used in Chapter 3 or in the proof of (5.4.12) given below; hence it will be omitted. Note that one can establish more than the $a.s.$ convergence property of $\widetilde{\theta}_t$ to θ^*. Indeed, from Theorem 3.4.2, one can straightforwardly show that under $E1$ and $E2$, we have

$$t^{1/2-\delta}\left(\widetilde{\theta}_t-\theta^*\right)\xrightarrow[t\to\infty]{a.s.}0, \tag{5.4.13}$$

for any $\delta>0$.

The proof of (5.4.12) is based on the two lemmas given below. Lemma 5.4.1 establishes a condition of the summability of the terms $t^{-1}h_t\widetilde{h}_t^{-2}\widehat{\varphi}_t'\widehat{R}_t^{-1}\widehat{\varphi}_t$, which is a crucial point in the convergence of $\widehat{\theta}_t$ to θ^*; while lemma 5.4.2 shows that this latter condition is guaranteed by assumption $E1$.

Lemma 5.4.1

Provided that the following condition

$$\lim_{t\longrightarrow\infty}\sup t^{-1}\sum_{k=1}^{t}\widetilde{h}_k^{-2}\left\|\widehat{\varphi}_k\right\|^2<\infty,\ a.s., \tag{5.4.14}$$

holds, we have for any $\delta>0$

$$\sum_{t=1}^{\infty}t^{-\delta}\widetilde{h}_t^{-2}\widehat{\varphi}_t'\widehat{R}_t^{-1}\widehat{\varphi}_t<\infty,\quad a.s., \tag{5.4.15}$$

where $\widetilde{h}_t = \widetilde{\varphi}_t' \widetilde{\theta}_{t-1}$.

Proof

Result (5.4.15a) follows from the properties of the inversion matrix lemma revealed by Solo (1979). ■

Now we need to show that $E1$ implies (5.4.14). This is given by the following lemma.

Lemma 5.4.2

If the assumption $E1$ *holds? which implies*

$$\lim_{t \longrightarrow \infty} \sup \frac{1}{t} \sum_{k=1}^{t} h_k^{-2} \varepsilon_k^4 < \infty, \tag{5.4.16}$$

then (5.4.14) *is satisfied.*

Proof

We need first to show that (5.4.16) implies

$$\lim_{t \longrightarrow \infty} \sup \frac{1}{t} \sum_{k=1}^{t} \widetilde{h}_k^{-2} \varepsilon_k^4 < \infty. \tag{5.4.17}$$

Thus, it is sufficient to show

$$S_t = \frac{1}{t} \sum_{k=1}^{t} \left((\widetilde{\varphi}_k' \widetilde{\theta}_k)^{-2} - (\varphi_k' \theta^*)^{-2} \right) \varepsilon_k^4 = O(1), \ a.s.$$

Using the mean value theorem as in Bose and Mukherjee (2003), we can rewrite S_t as follows

$$S_t = 2t^{-1} \sum_{k=1}^{t} \chi_k^{-3} \left(\widetilde{\varphi}_k' \widetilde{\theta}_k - \varphi_k' \theta^* \right) \varepsilon_k^4,$$

where from the inequality given in Bose and Mukherjee (2003, formula (17), p. 133) χ_k is such that

$$\begin{aligned} 0 \ &< \ \chi_k^3 < \left(\varphi_k' \theta^*\right)^{-3} \left[1 + \frac{\varphi_k' \theta^*}{\varphi_k' \widetilde{\theta}_k} \frac{\widetilde{\theta}_k' \varphi_k}{\widetilde{\theta}_k' \widetilde{\varphi}_k} \right]^3, \\ &< \ \left(\alpha_{k',0}^*\right)^{-3} \left[1 + \left(1 + \sum_{j=1}^{p+q+1} \frac{\theta_{k',j}^*}{\widetilde{\theta}_{k',j}^{(k)}} \right) \left(1 + q + \sum_{j=1}^{p} \frac{\varepsilon_{k-j}^2 - \upsilon_{k-j}}{\varepsilon_{k-j}^2 - \widetilde{\vartheta}_{k-j}} \right) \right]^3, \end{aligned} \tag{5.4.18}$$

k' is the remainder of the division of k by S and where the upper bound of the last inequality is bounded *a.s.* in view of (5.4.11). In addition, this last expression allows us to obtain the inequality

$$\begin{aligned}\left|\widetilde{\varphi}_k'\widetilde{\theta}_k - \varphi_k'\theta^*\right| &\leq \left|\widetilde{\varphi}_k'(\widetilde{\theta}_k - \theta^*)\right| + \left|(\widetilde{\varphi}_k - \varphi_k)'\,\theta^*\right|, \\ &\underset{k\to\infty}{\longrightarrow} 0,\ a.s.\end{aligned}$$

Thus, from the Toeplitz lemma and $E1$ we have

$$\begin{aligned}|S_t| &\leq 2t^{-1}\sum_{k=1}^{t}\chi_k^{-3}\left|\widetilde{\varphi}_k'\widetilde{\theta}_k - \varphi_k'\theta^*\right|\varepsilon_k^4, \\ &\underset{t\to\infty}{\longrightarrow} 0,\ a.s.,\end{aligned}$$

where C^* is the bounding constant in (5.4.18). In the same manner, we can prove that under $E1$, we also have

$$\lim_{t\longrightarrow\infty}\sup\frac{1}{t}\sum_{k=1}^{t}\widetilde{h}_k^{-2} < \infty. \tag{5.4.19}$$

Now, it remains to show that (5.4.17) implies

$$\lim_{t\longrightarrow\infty}\sup\frac{1}{t}\sum_{k=1}^{t}\widetilde{h}_k^{-2}\widetilde{\vartheta}_k^2 < \infty. \tag{5.4.20}$$

However, we can easily show that (see Solo (1979) or Ljung and Söderström (1983, Lemma 4.2))

$$\frac{1}{t}\sum_{k=1}^{t}\frac{\widetilde{h}_k^{-2}\widetilde{\vartheta}_k^2}{1-\widetilde{h}_k^{-2}\widehat{\varphi}_k'\widehat{R}_k^{-1}\widehat{\varphi}_k} = \frac{1}{t}\sum_{k=1}^{t}\widetilde{h}_k^{-2}\varepsilon_k^4 + \frac{1}{t}\widehat{\theta}_0'\widehat{R}_0\widehat{\theta}_0 - \frac{1}{t}\widehat{\theta}_t'\widehat{R}_t\widehat{\theta}_t. \tag{5.4.21}$$

Thus, from (5.4.21), it is clear that (5.4.20) gives

$$\lim_{t\longrightarrow\infty}\sup\frac{1}{t}\sum_{k=1}^{t}\frac{\widetilde{h}_k^{-2}\widetilde{\vartheta}_k^2}{1-\widetilde{h}_k^{-2}\widehat{\varphi}_k'\widehat{R}_k^{-1}\widehat{\varphi}_k} < \infty,$$

and as $\widetilde{h}_k^{-2}\widehat{\varphi}_k'\widehat{R}_k^{-1}\widehat{\varphi}_k < 1$ (cf, Appendix A) then (5.4.20) is true. From (5.4.19) we find that

$$\lim_{t\longrightarrow\infty}\sup\frac{1}{t}\sum_{k=1}^{t}\left[\widetilde{h}_k^{-2} + \sum_{i=1}^{q}\widetilde{h}_k^{-2}\varepsilon_{k-i}^4 + \sum_{j=1}^{p}\widetilde{h}_k^{-2}\left(\varepsilon_{k-j}^2 - \widetilde{\vartheta}_{k-j}\right)^2\right] < \infty$$

which is exactly (5.4.14). Hence the proof is achieved. ■

Let us return now to prove (5.4.12). Setting $\overline{\theta}_t = \widehat{\theta}_t - \theta^*$, $T_t = \overline{\theta}'_t \widehat{R}_t \overline{\theta}_t$, $a_t = -\widetilde{h}_t^{-1} \widehat{\varphi}'_t \overline{\theta}_t$, $b_t = \widetilde{h}_t^{-1}\left(\widetilde{\vartheta}_t - v_t - \widetilde{h}_t \frac{1}{2} a_t\right)$, and $\overline{T}_t = \frac{1}{t}\left(T_t + 2\sum_{k=1}^{t} \widetilde{h}_k^{-2} a_k b_k\right)$, then along the lines of Chapter 3, we can easily obtain the relation

$$\begin{aligned} t\overline{T}_t &= (t-1)\overline{T}_{t-1} + 2\widetilde{h}_t^{-2}\widehat{\varphi}'_t\widehat{R}_t^{-1}\widehat{\varphi}_t v_t^2 + 2\widetilde{h}_t^{-2}\widehat{\varphi}'_t\overline{\theta}_{t-1}v_t + \\ &\qquad 2h_t^{-2}\widehat{\varphi}'_t\widehat{R}_t^{-1}\widehat{\varphi}_t(\widehat{v}_t - v_t)v_t - \widetilde{h}_t^{-4}\widehat{\varphi}'_t\widehat{R}_{t-1}^{-1}\widehat{\varphi}_t\widehat{\vartheta}_t^2 \\ &\le (t-1)\overline{T}_{t-1} + 2\widetilde{h}_t^{-2}\widehat{\varphi}'_t\widehat{R}_t^{-1}\widehat{\varphi}_t v_t^2 + 2\widetilde{h}_t^{-2}\widehat{\varphi}'_t\overline{\theta}_{t-1}v_t + 2\widetilde{h}_t^{-2}\widehat{\varphi}'_t\widehat{R}_t^{-1}\widehat{\varphi}_t(\widehat{v}_t - v_t)v_t \end{aligned} \tag{5.4.22}$$

Since $E(v_t/\mathcal{F}_{t-1}) = 0$, $E(v_t^2/\mathcal{F}_{t-1}) = \left(E(\eta_t^4) - 1\right)h_t = Kh_t$ and the quantities : $\widetilde{h}_t$, $\widehat{\varphi}_t$, $\widehat{R}_t^{-1}$, $\overline{\theta}_{t-1}$, $\widehat{v}_t - v_t$ are $\mathcal{F}_{t-1}$-measurable, then taking the conditional expectation of both sides of (5.4.22) with respect to $\mathcal{F}_{t-1}$, we obtain

$$E(\overline{T}_t/\mathcal{F}_{t-1}) \le \overline{T}_{t-1} - t^{-1}\overline{T}_{t-1} + 2Kt^{-1}\widetilde{h}_t^{-2}\widehat{\varphi}'_t\widehat{R}_t^{-1}\widehat{\varphi}_t h_t \tag{5.4.23}$$

Now, we can apply Robbins and Siegmund's (1971) Martingale theorem (1971), to (5.4.23). We have only to show that $\overline{T}_{t-1}$ is positive and that the two last terms in (5.4.23) are summable. However, these last terms are effectively summable in view of Lemma 5.4.1, Lemma 5.4.2, assumption $E1$, the Markov inequality and the Borel-Cantelli lemma. Then it remains to show that

$$\sum_{k=1}^{t} \widetilde{h}_k^{-2} a_k b_k \ge 0 \tag{5.4.24}$$

Note that, from the definition of a_k and b_k, one can easily show that (See also Adams and Goodwin, 1995)

$$b_{s+S\tau} = ((1-\beta_s(L))^{-1} - 1/2)a_{s+S\tau} = H_s(L)a_{s+S\tau} \tag{5.4.25}$$

Then, (5.4.24) can be easily obtained as in Solo (1979), whenever $E2$ is verified. Indeed, we have

$$\sum_{k=1}^{t} a_k b_k \ge \lambda \sum_{k=1}^{t} a_k^2$$

where $\lambda = \inf_{-\pi\le\omega\le\pi, s\in\{1,..,S\}} \left[\mathrm{Re}(1-\beta_s(e^{i\omega}))^{-1} - 1/2\right]$ and $k' = k \bmod[S]$. Finally, application of the Robbins and Siegmund (1971) theorem to (5.4.23) gives (5.4.12). ■

It is worth noting that, when $\liminf \frac{1}{t}\widetilde{R}_t > 0$ and $\liminf \frac{1}{t}\widehat{R}_t > 0$, (cf, Solo, 1979), the results (5.4.11a) and (5.4.12a) imply

$$\widetilde{\theta}_t \xrightarrow{a.s.} \theta^*,$$

$$\widehat{\theta}_t \xrightarrow{a.s.} \theta^*.$$

Remark 5.4.1

It maybe possible to reduce the number of moment assumptions by normalisation of the regression function (5.4.5) to obtain a normalised recursive on-line pre-estimator. Such an estimator would be similar in spirit to Moulines et al. (2003) and Dahlhaus and Subba Rao (2005) (though for technical reasons it seems that Dahlhaus and Subba Rao (2005) assume existence of the 4th moment of the process). We note that the algorithms considered in the references given above are for recursive on-line parameter estimation of (nonstationary) AR or $ARCH$ processes - the normalisation in a recursive on-line $GARCH$ algorithm would have to be different. Of course one must remember if one normalises the algorithm, the spectral arguments that we used in $E2$ may not apply.

5.4.1 Summary

This section concentrates on the problem of recursive parameter estimation of $PGARCH$ models. In particular a pseudo linear regression method was proposed for estimating the $PGARCH$ parameters. The basic method of the algorithm was based on a two-stage scheme. First an on-line recursive pre-estimate was proposed which used the regression representation of the process :

$$\varepsilon_t^2 = \varphi_t'\theta + h_t\left(\zeta_t^2 - 1\right),$$

where we note that $\left\{h_t\left(\zeta_t^2 - 1\right)\right\}_t$ are Martingale differences. A type of recursive least squares was used to obtain estimates of the parameters. The pre-estimator was used to obtain the main estimator, where the above equation was normalised by an estimate of conditional variance h_t (say $\widetilde{h}_t = \widetilde{\varphi}_t'\widetilde{\theta}_{t-1}$) and a recursive on-line method was proposed based on the regression representation

$$\frac{\varepsilon_t^2}{\widetilde{h}_t} = \frac{\varphi_t'\theta}{\widetilde{h}_t} + \frac{h_t\left(\zeta_t^2 - 1\right)}{\widetilde{h}_t},$$

The estimator in the second stage was considered better than the pre-estimator, since in the limit the Martingale differences $\left\{\left(\zeta_t^2-1\right)\dfrac{h_t}{\widetilde{h}_t}\right\}_t$ should be now independent.

5.5 A two-stage recursive predictor error algorithm

The desirable strong convergence property of the obtained $TS-PLR$ estimators $\widetilde{\theta}_t$ and $\widehat{\theta}_t$ is ensured by Theorem 5.4.1; however, there is no, in our knowledge, available theoretical results concerning the asymptotic distribution of these estimators. Moreover, aside from modest simulation tools (see Ljung and Söderström, 1983) which can suggest the superiority of $\widehat{\theta}_t$ over $\widetilde{\theta}_t$, it appears that there is no existing results in the literature concerning the theoretical study of either the asymptotic efficiency of the two estimators $\widehat{\theta}_t$ and $\widetilde{\theta}_t$ nor their relative efficiency. Concerning these important points, Ljung and Söderström (1983, p. 231) showed via a particular counterexample, without giving the corresponding asymptotic distribution, that the PLR estimator is not asymptotically efficient. To avoid these inconveniences, we propose, as in Chapter 3, to replace the vectors $\widetilde{\varphi}_t$ and $\widehat{\varphi}_t$, in (5.4.8) by the vectors $\widetilde{\psi}_t$ and $\widehat{\psi}_t$ obtained through a linear filter as we show below. Doing this, we can establish a two-stage recursive estimation procedure which is an analogous to the well-known Recursive Maximum Likelihood (RML) method.

Letting $\psi_t(\theta)=\dfrac{\partial\widehat{\varepsilon}_t(\theta)}{\partial\theta'}$, where $\psi_t(\theta^*)=-\dfrac{\partial v_t}{\partial\theta'}$, and deriving with respect to θ the two sides of (3.4), we obtain

$$\psi_t(\theta)-\sum_{j=1}^{p}\beta_{s,j}\psi_{t-j}(\theta)=\varphi_t(\theta), \tag{5.5.1}$$

where $s=t \bmod [S]$ and $\varphi_t(\theta)$ is obtained from φ_t,given in (5.4.4), by replacing v_t by its approximate $v_t(\theta)$. Replacing in (5.5.1), $\psi_t(\theta)$ and $\varphi_t(\theta)$ by their approximates $\widetilde{\psi}_t$ and $\widetilde{\varphi}_t$, respectively ,and on a second time by their approximates $\widehat{\psi}_t$ and $\widehat{\varphi}_t$, respectively, we obtain the following relations :

$$\begin{aligned}\widetilde{\psi}_t-\sum_{j=1}^{p}\widetilde{\beta}_{s,j}^{(t-1)}\widetilde{\psi}_{t-j}&=\widetilde{\varphi}_t,\\ \widehat{\psi}_t-\sum_{j=1}^{p}\widehat{\beta}_{s,j}^{(t-1)}\widehat{\psi}_{t-j}&=\widehat{\varphi}_t.\end{aligned} \tag{5.5.2}$$

Using the previous formulation, we are able to build, as was already made for the RML algorithm, (see Chapter 3), the following algorithm.

Algorithm 5.5.1 ($TS-RPE$ algorithm) *The TS-RPE algorithm for estimating a PGARCH model is given by the following recursive equations*

$$\underbrace{\begin{cases} \widetilde{v}_t = \varepsilon_t^2 - \widetilde{\varphi}_t'\widetilde{\theta}_{t-1}, \\ \widetilde{\theta}_t = \widetilde{\theta}_{t-1} + \dfrac{\widetilde{P}_{t-1}\widetilde{\psi}_t\widetilde{v}_t}{1+\widetilde{\psi}_t'\widetilde{P}_{t-1}\widetilde{\psi}_t}, \ \widetilde{\theta}_0 = 0, \\ \widetilde{P}_t = \widetilde{P}_{t-1} - \dfrac{\widetilde{P}_{t-1}\widetilde{\psi}_t\widetilde{\psi}_t'\widetilde{P}_{t-1}}{1+\widetilde{\psi}_t'\widetilde{P}_{t-1}\widetilde{\psi}_t}, \ \widetilde{P}_0 = M.I, \\ \widetilde{\psi}_{t+1} = \sum\limits_{k=1}^{q} \widetilde{\beta}_{s,k}^{(t)}\widetilde{\psi}_{t-k+1} + \widetilde{\varphi}_{t+1}, \end{cases}}_{\text{first stage}} , \underbrace{\begin{cases} \widehat{v}_t = \varepsilon_t^2 - \widehat{\varphi}_t'\widehat{\theta}_{t-1}, \\ \widehat{\theta}_t = \widehat{\theta}_{t-1} + \dfrac{\widehat{P}_{t-1}\widehat{\psi}_t\widehat{v}_t}{\left(\widetilde{\varphi}_t'\widetilde{\theta}_{t-1}\right)^2 + \widehat{\psi}_t'\widehat{P}_{t-1}\widehat{\psi}_t}, \ \widehat{\theta}_0 = 0 \\ \widehat{P}_t = \widehat{P}_{t-1} - \dfrac{\widehat{P}_{t-1}\widehat{\psi}_t\widehat{\psi}_t'\widehat{P}_{t-1}}{\left(\widetilde{\varphi}_t'\widetilde{\theta}_{t-1}\right)^2 + \widehat{\psi}_t'\widehat{P}_{t-1}\widehat{\psi}_t}, \ \widehat{P}_0 = MI \\ \widehat{\psi}_{t+1} = \sum\limits_{k=1}^{q} \widehat{\beta}_{s,k}^{(t)}\widehat{\psi}_{t-k+1} + \widehat{\varphi}_{t+1}, \end{cases}}_{\text{second stage}} \tag{5.5.3}$$

where $s = (t+1) \bmod [S]$ and $\widetilde{\varphi}_t$ and $\widehat{\varphi}_t$ are given in Section 3.

It is worth noting that the consistency property of the estimators $\widetilde{\theta}_t$ and $\widehat{\theta}_t$ obtained by the recursive system (5.5.3) can be easily established as in (5.4.11) and (5.4.12) by only using $\widetilde{\psi}_t$ and $\widehat{\psi}_t$ in place of $\widetilde{\varphi}_t$ and $\widehat{\varphi}_t$, respectively, and by replacing the assumption $E2$ with the following assumptions :

$E3$: We suppose that Algorithm 5.5.1 includes a projection that keeps $\widetilde{\theta}_t$ and $\widehat{\theta}_t$ in the stability region.

$E4$:

$$\operatorname{Re}\left[(1-\beta_s(e^{i\omega}))^{-1}(1-\widetilde{\beta}_s^{(t-1)}(e^{i\omega})) - \frac{1}{2}\right] > 0 \qquad \forall\omega \in \left]-\pi,\pi\right[\tag{5.5.4a}$$

$$\operatorname{Re}\left[(1-\beta_s(e^{i\omega}))^{-1}(1-\widehat{\beta}_s^{(t-1)}(e^{i\omega})) - \frac{1}{2}\right] > 0 \qquad \forall\omega \in \left]-\pi,\pi\right[. \tag{5.5.4b}$$

The latter assumptions hold whenever $\beta_s(L)$ is stable for all $s = 1, ..., S$, where $\widetilde{\beta}_s^{(t)}(L) = \sum\limits_{j=1}^{p} \widetilde{\beta}_{s,j}^{(t)} L^j$ and $\widehat{\beta}_s^{(t)}(L) = \sum\limits_{j=1}^{p} \widehat{\beta}_{s,j}^{(t)} L^j$.

In addition to the establishment of the strong convergence property of the $TS-RPE$ estimators, we obtain under the following additional assumptions, the asymptotic distribution property of these estimators.

$E5$: There exists $M^* > 0$ such that $\left|\varepsilon_t^2\right| < M^*$, *a.s.*

Letting $F_1 = \sum\limits_{s=1}^{S} E\left(\psi_s(\theta^*)\psi_s'(\theta^*)\right)$, $F_2 = \sum\limits_{s=1}^{S} E\left(\psi_s(\theta^*)\psi_s'(\theta^*)v_s^2\right)$ and $F_3 = \sum\limits_{s=1}^{S} E\left(\psi_s(\theta^*)\psi_s'(\theta^*)h_s^{-2}\left(\theta^*\right)\right)$, we can state the following results.

Theorem 5.4.2 *Under the conditions $E1 - E5$, we have*

$$t^{1/2}(\widetilde{\theta}_t - \theta^*) \xrightarrow{D} N\left(0, SF_1^{-1}F_2F_1^{-1}\right), \tag{5.5.5}$$

and

$$t^{1/2}(\widehat{\theta}_t - \theta^*) \xrightarrow{D} N\left(0, SVar(\eta_s^2)F_3^{-1}\right). \tag{5.5.6}$$

where "$\xrightarrow{D}$" stands, as usual, for the convergence in law.

Proof

The proof of results (5.5.5) and (5.5.6) can be easily drawn by similar techniques as those used for other estimation algorithms stated in Solo (1981), Ljung and Söderström (1983, Appendix 4.B) or in Chapter 3. Hence, we only give a sketch of the proof of result (5.5.6) while the proof of (5.5.5) can be drawn similarly. By backward recursion, one can obtain from the last three recursion relations in (5.5.3), the following relation

$$\widehat{R}_t\overline{\theta}_t = \widehat{R}_0\overline{\theta}_0 + \sum_{k=1}^{t}\widehat{\psi}_k\frac{v_k}{\left(\widetilde{\varphi}_k'\widetilde{\theta}_k\right)^2} + \sum_{k=1}^{t}\widehat{\psi}_k\left(\frac{\widehat{\psi}_k'\overline{\theta}_{k-1} + \widehat{v}_k - v_k}{\left(\widetilde{\varphi}_k'\widetilde{\theta}_k\right)^2}\right). \tag{5.5.7}$$

Now we analyze each term of (5.5.7) which can be written as

$$t^{1/2}\overline{\theta}_t = \left(\frac{1}{t}\widehat{R}_t\right)^{-1} t^{-1/2}\sum_{k=1}^{t}\psi_k\left(\theta^*\right)\frac{v_k}{\left(\widetilde{\varphi}_k'\widetilde{\theta}_k\right)^2} + t^{-1/2}\left(\frac{1}{t}\widehat{R}_t\right)^{-1}\omega_t, \tag{5.5.8}$$

where

$$\omega_t = \sum_{k=1}^{t}\left(\widehat{\psi}_k - \psi_k\left(\theta^*\right)\right)\frac{v_k}{\left(\widetilde{\varphi}_k'\widetilde{\theta}_k\right)^2} + \widehat{R}_0\overline{\theta}_0 + \sum_{k=1}^{t}\widehat{\psi}_k\left(\frac{\widehat{\psi}_k'\overline{\theta}_{k-1} + \widehat{v}_k - v_k}{\left(\widetilde{\varphi}_k'\widetilde{\theta}_k\right)^2}\right). \tag{5.5.9}$$

First, we have to show that

$$\frac{1}{t}\widehat{R}_t \xrightarrow{a.s} \frac{1}{S}\sum_{s=1}^{S} E\left(\frac{\psi_s(\theta^*)\,\psi_s'(\theta^*)}{(\varphi_s'\theta^*)^2}\right), \text{ as } t \longrightarrow \infty. \tag{5.5.10}$$

However,

$$\begin{aligned}\frac{1}{t}\widehat{R}_t &= \frac{1}{t}\sum_{k=1}^{t}\frac{\widehat{\psi}_k\widehat{\psi}_k'}{\left(\widetilde{\varphi}_k'\widetilde{\theta}_k\right)^2},\\ &= \frac{1}{t}\sum_{k=1}^{t}\frac{\widehat{\psi}_k\widehat{\psi}_k'}{(\varphi_k'\theta^*)^2} + \frac{1}{t}\sum_{k=1}^{t}\left(\frac{1}{\left(\widetilde{\varphi}_k'\widetilde{\theta}_k\right)^2} - \frac{1}{(\varphi_k'\theta^*)^2}\right)\widehat{\psi}_k\widehat{\psi}_k'.\end{aligned}$$

From $E3$ and $E4$, it is easy to show that the first term can by written as

$$\begin{aligned}\frac{1}{t}\sum_{k=1}^{t}\frac{\widehat{\psi}_k\widehat{\psi}_k'}{(\varphi_k'\theta^*)^2} &= \frac{1}{t}\sum_{k=1}^{t}\frac{\psi_k(\theta^*)\,\psi_k'(\theta^*)}{(\varphi_k'\theta^*)^2} + \frac{1}{t}\sum_{k=1}^{t}\frac{\left(\widehat{\psi}_k - \psi_k(\theta^*)\right)\widehat{\psi}_k'}{(\varphi_k'\theta^*)^2}\\ &\quad + \frac{1}{t}\sum_{k=1}^{t}\frac{\left(\widehat{\psi}_k - \psi_k(\theta^*)\right)\psi_k'(\theta^*)}{(\varphi_k'\theta^*)^2},\\ &= \frac{1}{t}\sum_{k=1}^{t}\frac{\psi_k(\theta^*)\,\psi_k'(\theta^*)}{(\varphi_k'\theta^*)^2} + o_p(1).\end{aligned}$$

and by the periodic stationarity and ergodicity, we obtain for $t = s + S\tau$

$$\begin{aligned}\frac{1}{t}\sum_{k=1}^{t}\frac{\widehat{\psi}_k\widehat{\psi}_k'}{(\varphi_k'\theta^*)^2} &= \frac{1}{S}\sum_{s=1}^{S}\frac{1}{\tau}\sum_{m=0}^{\tau-1}\frac{\psi_{s+Sm}(\theta^*)\,\psi_{s+Sm}'(\theta^*)}{\left(\varphi_{s+Sm}'\theta^*\right)^2} + o_p(1),\\ &\xrightarrow{a.s} \frac{1}{S}\sum_{s=1}^{S} E\left(\frac{\psi_s(\theta^*)\,\psi_s'(\theta^*)}{(\varphi_s'\theta^*)^2}\right).\end{aligned}$$

Hence (5.5.10) is established provided that we have

$$\frac{1}{t}\sum_{k=1}^{t}\left(\frac{1}{\left(\widetilde{\varphi}_k'\widetilde{\theta}_k\right)^2} - \frac{1}{(\varphi_k'\theta^*)^2}\right)\widehat{\psi}_k\widehat{\psi}_k' = o_p(1),$$

which can be obtained easily, along the lines of Lemma 5.4.2, by using again the mean value theorem, assumption $E4$ and the Toeplitz lemma. Considering the last term in the second member of (5.5.8), one can show, under $E3$ and $E4$ and according to Ljung and Söderström (1983, Lemmas 4.B1, 4.B4 and 4.B.6), that

$$t^{-1/2}\left(\frac{1}{t}\widehat{R}_t\right)^{-1}\omega_t = o_p(1).$$

Finally, to obtain (5.5.6), we need to establish the following asymptotic convergence in law

$$t^{-1/2}\sum_{k=1}^{t}\psi_k\left(\theta^*\right)\frac{v_k}{\left(\widetilde{\varphi}_k'\widetilde{\theta}_k\right)^2}\xrightarrow{D} N\left(0,\frac{1}{S}\sum_{s=1}^{S}E\left(\frac{\psi_s\left(\theta^*\right)\psi_s'\left(\theta^*\right)}{\left(\varphi_s'\theta^*\right)^2}\eta_s^2\right)\right).$$

But

$$\begin{aligned} t^{-1/2}\sum_{k=1}^{t}\psi_k\left(\theta^*\right)\frac{v_k}{\left(\widetilde{\varphi}_k'\widetilde{\theta}_k\right)^2} = {} & t^{-1/2}\sum_{k=1}^{t}\psi_k\left(\theta^*\right)\frac{v_k}{\left(\varphi_k'\theta^*\right)^2}+ \\ & t^{-1/2}\sum_{k=1}^{t}\left(\frac{1}{\left(\widetilde{\varphi}_k'\widetilde{\theta}_k\right)^2}-\frac{1}{\left(\varphi_k'\theta^*\right)^2}\right)\psi_k\left(\theta^*\right)v_k \end{aligned} \tag{5.5.11}$$

in which the last term of the second member of (5.5.11) is "$o_p(1)$". Indeed, by means of the mean values theorem we have

$$\begin{aligned} t^{-1/2+\delta}\sum_{k=1}^{t}\left(\frac{1}{\left(\widetilde{\varphi}_k'\widetilde{\theta}_k\right)^2}-\frac{1}{\left(\varphi_s'\theta^*\right)^2}\right)\psi_k\left(\theta^*\right)v_k &= \\ t^{-1/2+\delta}\sum_{k=1}^{t}\left(\frac{\widetilde{\varphi}_k'\widetilde{\theta}_k-\varphi_s'\theta^*}{\chi_k^3}\right)\psi_k\left(\theta^*\right)v_k &= S_t \end{aligned}$$

for some δ such that $0<\delta<1/2$, where as in Lemma B.2, $\dfrac{1}{\chi_k^3}$ is bounded almost surely. Since under $E1-E4$, the expectation $E\left\|S_t\right\|^2$ is finite for all t, it follows from the Tchebyshev inequality that $t^{-\delta}\left\|S_t\right\| = o_p(1)$ and hence $t^{-\delta}S_t = o_p(1)$. Thus, by the Martingale central limit theorem (cf, Hall and Heyde, 1980, Corollary 3.1) which is easily applicable and the periodic ergodicity and stationary of the underlying process we have

$$t^{-1/2}\sum_{k=1}^{t}\psi_k\left(\theta^*\right)\frac{v_k}{\left(\varphi_s'\theta^*\right)^2}\xrightarrow{D} N\left(0,\frac{1}{S}\sum_{s=1}^{S}E\left(\frac{\psi_s\left(\theta^*\right)\psi_s'\left(\theta^*\right)}{\left(\varphi_s'\theta^*\right)^2}\eta_s^2\right)\right),$$

which complete the proof.

5.6 Discussion

Although the algorithms presented above provide consistent estimates without the need for monitoring procedure, because of the positivity constraints on the parameters which cannot be assured by Newton-Raphson's nature, they must include a procedure which monitors the positivity and hence the stability of the estimates, in order to lie inside the positivity region. To take into account this important point, one can apply the monitoring procedure proposed by Ljung and Söderström (1983). The stability conditions can be checked, at each iteration, by the use of the results given in Chapter 4. However, the above monitoring procedure disturbs the rapidity of the algorithm. Moreover, the use of the constrained least squares estimator, for higher order models, seems to be very useful (see Pantula 1988; Wang et al., 2004). Nevertheless, these algorithms which can be applied even to off-line data, can take into account the currently dynamic character of many economic time series particularly the financial ones. Thus, these methods derived from the adaptive control field, will probably see a resurgence of real interest in the field of econometrics.

5.7 Simulation study

5.7.1 Case of $GARCH$ models

In this subsection we will test the performance of the aforementioned recursive algorithms via simulated data. Table 1 compares the results obtained by our TS-RLS estimator and those obtained by the algorithm of Bose and Mukherjee (2003) (noted by $B\&M$), on several simulated time series, generated from the standard Gaussian $ARCH(1)$ model with $(\alpha_0, \alpha_1) = (0.1, 0.3)$. For all simulations, we consider one thousand Monte Carlo replications. The initial parameter values are taken to be zeros and $\widetilde{P}_0 = \widehat{P}_0 = 100000I$. The mean ($M$) of the obtained parameter estimations of both algorithms and their corresponding mean square error (MSE) are reported, in the same table wich follows :

Data length	$\widehat{\alpha}_0^{(TS\text{-}RLS)}$ $M.$ MSE	$\widehat{\alpha}_1^{(TS\text{-}RLS)}$ $M.$ MSE	$\widehat{\alpha}_0^{(B\&M)}$ $M.$ MSE	$\widehat{\alpha}_1^{(B\&M)}$ $M.$ MSE
40	.0902 .0112	.2923 .0737	.1066 .0103	.2831 .0608
60	.0912 .0092	.3054 .0493	.1039 .0084	.2869 .0465
80	.0942 .0085	.2953 .0409	.1037 .0072	.2891 .0399
100	.0946 .0047	.2974 .0369	.1024 .0043	.2900 .0300

Observe that the TS-RLS estimates are good compared to the $B\&M$ estimates, even for small data length. For moderate data the TS-RLS performs better than its homologue because of the good monitoring of the positivity constraint.

In Table 2, we assess the performances of the TS-PLR and RPE estimates, based on simulated data generated from the Gaussian $GARCH(1,1)$ model with parameters $(\alpha_0, \alpha_1, \beta_1) = (.3, .05, .8)$. The starting values are taken as being the same as in the previous example.

Data length	$\widehat{\alpha}_0^{(PL)}$ $M.$ MSE	$\widehat{\alpha}_1^{(PL)}$ $M.$ MSE	$\widehat{\beta}_1^{(PL)}$ $M.$ MSE	$\widehat{\alpha}_0^{(RM)}$ $M.$ MSE	$\widehat{\alpha}_1^{(RM)}$ $M.$ MSE	$\widehat{\beta}_1^{(RM)}$ $M.$ MSE
80	.3371 .0925	.0622 .0081	.8541 .1028	.3445 .1079	.0653 .0091	.8465 .0912
300	.3213 .0649	.0588 .0066	.8341 .0897	.3208 .0592	.0547 .0063	.8265 .0714
600	.3105 .0417	.0518 .0047	.8241 .0718	.3086 .0491	.0509 .0043	.8165 .0676

Note that for small data length, the estimates given by our two last algorithms are acceptable and that the TS-PLR performs better than its homologue. However, for moderate and long times series, the RPE gives better estimates than the TS-PLR algorithm. We also note that the proposed algorithms must be monitored to cure the negative values in the obtained estimates, which violate the stability constraint. Indeed, these algorithms are based on the Newton-Raphon method which leads us to obtain some negative solutions. The monitoring procedure is derived from Ljung and Söderström (1983).

5.7.2 Case of periodic ARCH models

In this subsection we test for the performance of the above methods via simulated data. Indeed, their performance was assessed on several time series generated from periodic $ARCH$ of different orders and periods for a variety of sizes. However, for the sake of brevity, we consider only the results associated with simulated series of sizes, 80, 120, 160 and 200, from the standard Gaussian $PARCH_2(1)$ model with $(\alpha_{10}, \alpha_{11}, \alpha_{20}, \alpha_{21}) = (.05, .7, .3, .1)$. For each size, we consider one thousand Monte Carlo replications. The initial parameter values are taken to be zeros and $\widehat{P}_0 = \widetilde{P}_0 = 100000I$. The following table presents the results obtained by the $2S$-LS method (noted LS) and the $2S$-RLS algorithm (noted RS). M and MSE denote the mean of the obtained parameter estimations of both algorithms and their corresponding mean squares errors, respectively.

T	$\widehat{\alpha}_{10}^{(LS)}$ M MSE	$\widehat{\alpha}_{11}^{(LS)}$ M MSE	$\widehat{\alpha}_{20}^{(LS)}$ M MSE	$\widehat{\alpha}_{21}^{(LS)}$ M MSE	$\widehat{\alpha}_{10}^{(RS)}$ M MSE	$\widehat{\alpha}_{11}^{(RS)}$ M MSE	$\widehat{\alpha}_{20}^{(RS)}$ M MSE	$\widehat{\alpha}_{21}^{(RS)}$ M MSE
80	.0628 .0016	.6721 .1932	.2847 .0056	.1611 .0568	.0528 .0013	.6696 .1627	.2603 .0399	.1471 .0468
120	.0606 .0012	.6799 .1450	.2910 .0040	.1410 .0232	.0519 .0011	.6781 .1341	.2666 .0329	.1384 .0284
160	.0573 .0010	.6848 .1275	.2940 .0033	.1389 .0179	.0512 .0009	.6846 .1276	.2709 .0283	.1318 .0212
200	.0572 .0009	.7127 .1110	.2967 .0022	.1203 .0115	.0505 .0004	.6986 .1067	.2765 .0196	.1201 .0111

Observe that the $2S$-RLS estimates are good and are comparable to the off-line estimates $2S$-LS, even for small time series. Moreover, the second algorithm can take account of the good monitoring of the positivity and the periodically stationary conditions. However, the above monitoring procedure slows down the rapidity of the algorithm because of its Newton-Raphson's nature. Consequently, the use of the constrained least squares estimator seems to be very useful (see also Pantula, 1988 and Wang et al., 2004).

5.8 Conclusion

In this chapter we have proposed some recursive algorithms for estimating nonlinear heteroskedastic models independently of the restrictive normality assumption. These methods, derived from the on-line estimation, extend the recursive least squares method, the pseudo linear regression method and the recursive maximum likelihood method to the case of periodic $GARCH$ models and give, under general conditions, consistent estimators. The proposed methods can also be applied to off-line data for which they guarantee a good monitoring of estimates, in order to lie inside the stability and positivity regions. Moreover, they can be easily adapted to more general models such as $PARMAX$ models with $PGARCH$ errors. Thus, the aforementioned algorithms are recommended as well for their simplicity and precision as for their dynamic character, which is adequate for real-time time series such as continuous time financial models for which $GARCH$ models are good approximations.

Part III

Appendix

Chapter 6

Appendix A

The matrix inversion lemma (The Wooldbury Formula)

Lemma A.1 *Let A, B and C be three matrices of compatible dimensions so that the sum $A + BCD$ is possible.*

$$[A + BCD]^{-1} = A^{-1} - A^{-1}B\left[DA^{-1}B + C^{-1}\right]^{-1}DA^{-1}. \tag{A1}$$

Consequences :

Let R_t be a matrix given by

$$\begin{aligned} R_t &= \sum_{k=1}^{t} \lambda_k \xi_k \xi_k', \\ &= R_{t-1} + \lambda_t \xi_t \xi_t', \end{aligned}$$

where $\lambda_t > 0$. Then application of Lemma A.1 gives

$$\begin{aligned} R_t^{-1} &= \left(R_{t-1} + \lambda_t \xi_t \xi_t'\right)^{-1}, \\ &= R_{t-1}^{-1} - \frac{R_{t-1}^{-1}\xi_t\xi_t' R_{t-1}^{-1}}{\lambda_t^{-1} + \xi_t' R_{t-1}^{-1}\xi_t}. \end{aligned} \tag{A2}$$

This implies

$$\begin{aligned} \lambda_t R_t^{-1}\xi_t &= \lambda_t R_{t-1}^{-1}\xi_t - \frac{\lambda_t R_{t-1}^{-1}\xi_t\xi_t' R_{t-1}^{-1}\xi_t}{\lambda_t^{-1} + \xi_t' R_{t-1}^{-1}\xi_t}, \\ &= \frac{R_{t-1}^{-1}\xi_t}{\lambda_t^{-1} + \xi_t' R_{t-1}^{-1}\xi_t}, \end{aligned} \tag{A3}$$

and then

$$\begin{aligned} \lambda_t \xi_t' R_t^{-1} \xi_t &= \frac{\xi_t' R_{t-1}^{-1} \xi_t}{\lambda_t^{-1} + \xi_t' R_{t-1}^{-1} \xi_t} \\ &< 1 \end{aligned} \tag{A4}$$

Applying again the forgoing lemma, we obtain

$$\left(\lambda_t^{-1} + \xi_t' R_{t-1}^{-1} \xi_t\right)^{-1} = \lambda_t \left(1 - \lambda_t \xi_t' R_{t-1}^{-1} \xi_t\right) \tag{A5}$$

and

$$\begin{aligned} R_{t-1}^{-1} &= \left(R_t - \lambda_t \xi_t \xi_t'\right)^{-1}, \\ &= R_t^{-1} - \frac{R_t^{-1} \xi_t \xi_t' R_t^{-1}}{-\lambda_t^{-1} + \xi_t' R_t^{-1} \xi_t}. \end{aligned} \tag{A6}$$

Therefore

$$\lambda_t R_{t-1}^{-1} \xi_t = \frac{R_t^{-1} \xi_t}{\lambda_t^{-1} - \xi_t' R_t^{-1} \xi_t}. \tag{A7}$$

Chapter 7

Appendix B

1- **Some usual inequalities**

We restate in this appendix without comment a few of the most important inequalities used earlier.

Let X and Y be r.v. on a probability space (Ω, A, P). Then :

C_r**-inequality :** $E|X+Y|^r \leq C_r \{E|X|^r + E|Y|^r\}$ for $r>0$, $C_r \equiv 2^{(r\vee 1)-1}$.

Hölder : $E|XY| \leq (E|X|^r)^{1/r} (E|Y|^s)^{1/s}$ for $r>1$, $\frac{1}{r}+\frac{1}{s}=1$.

Markov : $P(|X| \geq \eta) \leq \dfrac{E|X|^r}{\eta^r}$ for all $\eta > 0$; requires $r > 0$.

Dispersion : $E|X| \leq E|X+Y|$ for independent r.v.s with 0 means.

Jensen : $g(EX) \leq Eg(X)$ if g is convex on some $(a,b) \subset \mathbb{R}$ having $P(X \in (a,b)) = 1$, and if $E(X)$ is finite.

2- **Some usual Lemmas**

Toeplitz Lemma : Let a_{nk} (for $1 \leq k \leq k_n$, with $k_n \longrightarrow \infty$) be such that :

(i) For every fixed k, we have $a_{nk} \to 0$.

(ii) $\sum_{k=1}^{k_n} |a_{nk}| < c < \infty$, for every n.

Let $x'_n = \sum_{k=1}^{k_n} a_{nk} x_k$. Then

(a) $x_n \longrightarrow 0$ implies $x'_n \longrightarrow 0$.

If $\sum_{k=1}^{k_n} a_{nk} \to 1$, then

(b) $x_n \longrightarrow x$ implies $x'_n \longrightarrow x$

In particular, if $b_n \equiv \sum_{k=1}^n a_k \nearrow \infty$, then

(c) $x_n \longrightarrow x$ finite entails $\sum_{k=1}^n a_k x_k / b_n \to x$.

As a particular case :

Cesaro summability lemma : If $S_n = \sum_{k=1}^n x_k \to r$, then $\sum_{k=1}^n S_k/n \to r$.

Kronecker's lemma : Let $x_1, x_2, ...$ be arbitrary real numbers. If

$b_n \geq 0$ and $\nearrow \infty$, then

$\sum_1^n x_k \to r$ implies $\sum_1^n b_k x_k / b_n \to 0$.

Borel–Cantelli lemma :

We only state an important consequence of the Borel-Cantelli lemma due to Stout (1974).

Lemma (Stout, 1974, Theorem 2.1.1)

Let $(S_n)_{n\geq 1}$ *a sequence of r.v. such that*

$$\sum_{n\geq 1} P\left(|S_n| > \varepsilon\right) < \infty,$$

for every $\varepsilon > 0$. *Then*

$$S_n \overset{n\to\infty}{\to} 0, \ a.s.$$

Bibliography

[1] Abraham, A. and Ikenberry, D. L. (1994). The individual investor and the weekend effect. *Journal of Financial and Quantitative Analysis*, **29**, 263-78.

[2] Adams, G. J. and Goodwin, G. C. (1995). Parameter estimation for periodic $ARMA$ Models. *Journal of Time Series Analysis*, **16**, 127-145.

[3] Aknouche, A., (2006). Causality conditions and autocovariance calculations in $PVAR$ models. To appear in *Journal of Statistical Computation and Simulation.*

[4] Aknouche, A. and Bentarzi, M. (2005). On the existence of Higher-order moments of periodic $GARCH$ models. Submitted to *Statistics and Probability Letters.*

[5] Aknouche A. and Guerbyenne, H. (2005). On some probabilistic properties of double periodic AR models. Submitted to *Statistics and Probability Letters.*

[6] Aknouche, A. and Guerbyenne, H. (2006). Recursive estimation of $GARCH$ models. To appear in *Communications in Statistics- Simulation and Computation.*

[7] Aknouche, A. and Hamdi, F. (2006). Autocovariances and likelihood calculations for periodic $VARMA$ models. Submitted to *Computational Statistics and Data Analysis.*

[8] Aknouche, A. and Hamdi, F. (2006). Periodic Chandrasekhar recursions. Preprint.

[9] Andel, J. (1983). Statistical analysis of periodic autoregression. *Aplikace Matematiky*, **28**, 364-385.

[10] Anderson, P.L. and Meerschaert, M. M. (2005). Parameter Estimation for Periodically Stationary Time Series. *Journal of Time Series Analysis*, **26**, 489-518.

[11] Anderson, P.L., Meerschaert, M. M. and Vecchia, A.V. (1999). Innovations algorithm for periodically stationary time series. *Stochastic Processes and their Applications*, **83**, 149-169.

[12] Anderson, O. D. and Moore, J. B. (1979). *Optimal Filtering*. Prentice-Hall, INC, Englewood Cliffs, London.

[13] Ansley, C. F. (1979). An algorithm for the exact likelihood of a mixed autoregressive-moving average process. *Biometrika*, **66**, 59-65.

[14] Azrak, R. and Mélard, G. (1998). The exact quasi-likelihood of time-dependent ARMA models. *Journal of Statistical Planning and Inference*, **68**, 31-46

[15] Barnett, S. (1974). Simplification of the Lyapunov matrix equation $A^T PA - P = P$. *IEEE. Trans. Autom. Control*, **19**, 446-447.

[16] Barraud, A. Y. (1977). Numerical algorithms to solve $A^T XA - X = Q$. *IEEE. Trans. Autom. Control*, **22**, 883-885.

[17] Basawa, I. V. and Lund, R. (2001). Large sample properties of parameters estimates for periodic $ARMA$ models. Journal of Time Series Analysis, **22**, 1-13.

[18] Basawa, I. V., Lund, R. and Shao, Q. (2004). First-order seasonal autoregressive processes with periodically varying parameters. *Statistics and Probability Letters*, **67**, 299-306.

[19] Belister, E. (2000). Recursive estimation of a drifted autoregressive parameter. *The Annals of Statistics*, **28**, 860-870.

[20] 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.

[21] Bentarzi, M., (1998). Model-building problem of periodically m-variate moving average processes. *Journal of Multivariate Analysis*, **66**, 1-21.

[22] Bentarzi, M. and Aknouche A. (2004). Efficient two-stage estimation of periodic $ARCH$ models. Submitted to *The Econometrics Journal*.

[23] Bentarzi, M. and Aknouche A. (2005). Calculation of the Fisher information matrix for periodic $ARMA$ models. *Communications in Statistics-Theory and Methods*, **34**, 891-903.

[24] Bentarzi, M. and Aknouche A. (2006). An on-line estimation algorithm for periodic autoregressive models. To appear in *Communications in Statistics-Theory and Methods*, Vol **35**, issue 8.

[25] Bentarzi, M. and Hallin, M. (1994). On the Invertibility of Periodic Moving Average Models. *Journal of Time Series Analysis*, **15**, 263-268.

[26] Bentarzi, M. and Hallin, M. (1996). Locally optimal tests against periodic autoregression. *Econometric Theory*, **12**, 88-112.

[27] Bentarzi, M. and Hallin, M. (1998). Spectral factorization of periodically correlated $MA(1)$ processes. *Journal of Applied Probability*, **35**, 46-54.

[28] Benveniste, A., Métivier, M. and Priouret, P. (1990). *Adaptive algorithms and stochastic approximation*. Springer-Verlag, Berlin.

[29] Bera, A. K., Higgins, M. L. and Lee, S. (1992). Interaction between autocorrelation and conditional conditional heteroskedasticity : a random coefficient approach. *Journal of Business and Economic Statistics*, **10**, 133-142.

[30] Bera, A. K., Higgins, M. L. and Lee, S. (1996). Random coeffcients formulation of conditional heteroskedasticity and augmented $ARCH$ models. *Sankhya*, **58**, 199-220.

[31] Berkes, I., and Horvath, L. (2004). The efficiency of the estimators of the parameters in $GARCH$ processes. *Annals of statistics*, **32**, 633-655.

[32] Berkes, I., Horvath, L. and Kokoskza, P. (2003). $GARCH$ processes: Structure and estimation. *Bernoulli*, **9**, 201-227.

[33] Bessembinder, H. and Hertzel, M. G. (1993). Return autocorrelations around nontrading days. *Review of Financial Studies*, **6**, 155-89.

[34] Bhuiya, R. K. (1971). Stochastic analysis of periodic hydrologic processes. *Journal of Hydraulic Division of the American Society of Civil Engineering*, **97**, 949-962.

[35] Bibi, A. (2004). Probabilistic properties of periodic $GARCH$ processes. Preprint.

[36] Bibi, A. and Francq, C. (2003). Consistent and asymptotically normal estimators for time-dependent linear models. *Annals of the Institute of Statistical Mathematics*, **55**, 41-68.

[37] Bibi, A. and Gauthier, A. (2006). Propriétés dans L^2 et estimation des processus purement bilinéaires et strictement superdiagonaux à coefficients périodiques. *Forthcoming in The Canadian Journal of Statistics*, **34**.

[38] Bierman, G. J. (1977). *Factorization methods for discrete sequential estimation*. Academic Press, New York.

[39] Billingsley, P. (1968). *Convergence of probability measure*. Wiley, New York.

[40] Bittanti, S. Colaneri, P. and De Nicolao G. D. (1988). The difference periodic Riccati equation for the periodic prediction problem. *IEEE Trans. Autom. Control*, **33**, 706-712.

[41] Bittanti, S. and De Nicolao, G. (1993). Spectral factorization of linear periodic systems with application to the optimal prediction of periodic $ARMA$ models. *Automatica*, **29**, 517-522.

[42] Bojanczyk, A.W., Golub, G. and Van Dooren, P. (1992). The periodic Schur decomposition : Algorithms and applications, In F.T. Luk (ed.). *Proceedings SPIE Conference*, **1770**, 31-42.

[43] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **31**, 307-327.

[44] Bollerslev, T. and Ghysels, E. (1996). Periodic Autoregressive Conditional Heteroskedasticity. *Journal of Business and Economic Statistics*, **14**, 139-152.

[45] Bollerslev, T., Chou, R. Y. and Kroner, K. F. (1992). $ARCH$ modelling in finance. *Journal of Econometrics*, **52**, 5-59.

[46] Bordignon, S. and Lisi, F. (2000). Nonlinear analysis and prediction of river flow time series. *Evironmetrics*, **11**, 463-477.

[47] Bose, A. and Mukherjee, K. (2003). Estimating the $ARCH$ parameters by solving linear equations. *Journal of Time Series Analysis*, **24**, 127-136.

[48] Boshnakov, G. N. (1996). Recursive computation of the parameters of periodic autoregressive moving-average processes. *Journal of Time Series Analysis*, **17**, 333-349.

[49] Boshnakov, G. N. (1997). Periodically correlated solutions to a class of stochastic difference equations. *In: Csiszr, I., Michaletzky, G. Y. (Eds.), Stochastic differential and difference equations*. Birkhauser, Basel, 1-9.

[50] Boteva, A. and Boshnakov, G. N. (1992). An algorithm for the computation of the theoretical autocovariances of a periodic autoregressive process, *In Proceedings of the 8th Seminar on Statistical Data Analysis. Varna*, 9-15.

[51] Bougerol, P. and Picard, N. (1992*a*). Stationarity of $GARCH$ processes and some nonnegative time series. *Journal of Econometrics*, **52**, 115-127.

[52] Bougerol, P. and Picard, N. (1992*b*). Strict stationarity of generalised autoregressive processes. *Annals of Probability*, **20**, 1714-1730.

[53] Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis*, *Forecasting and Control*, 2nd edi., Holden-Day San Francisco, CA.

[54] Boyles, R. A. and Gardner, W. A. (1983). Cycloergodic properties of discrete-parameter nonstationary stochastic processes. *IEEE, trans. Infor. Theory.*, **29**, 105-114.

[55] Brandt, A. (1986). The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Adv. in Appl. Probab.*, **18**, 211-220.

[56] Brockwell, P. J. and Davis, R. A. (1991). *Time Series* : *Theory and methods*. Springer-Verlag. New York.

[57] Chen, M., An, H. Z. (1998). A note on the stationarity and the existence of moments of the $GARCH$ model. *Statistica Sinica*, **8**, 505-510.

[58] Čipra, T. (1985). Periodic moving average processes. *Aplikace Matematiky*, **30**, 218-229.

[59] Cleveland, W. P and Tiao, G. C. (1979). Modeling seasonal time series. *Revue Economique Appliquée*, **32**, 107-129.

[60] Cramér, H. (1961). On some classes of nonstationary stochastic processes. *Proceeding of the 4th Berkley Symposium on Mathematical Statistics and Probability*, **2**, 57-78.

[61] Dahlhauss, R. (1997). Fitting time series to nonstationary processes. *The Annals of Statistics*, **25**, 1-37.

[62] Dahlhaus, R. and Subba Rao, S. (2005). A recursive online algorithm for the estimation of time-varying $ARCH$ parameters. Preprint.

[63] Dahlhaus, R. and Subba Rao, S. (2006). Statistical inference for time-varying $ARCH$ processes. To appear in *The Annals of statistics*, Vol. **34**.

[64] De Jong, P. and Penzer, J. (2004). The $ARMA$ model in state space form. *Statistics and Probability Letter*, **70**, 119-125.

[65] Dent, W. (1977). Computation of the exact likelihood function of an $ARIMA$ process. *Journal of Statistical Computation and Simulation*, **5**, 193-206.

[66] Djuric, P. M. and Kay, M. (1992). Order selection of autoregressive models. *IEEE Trans. Signal. process.*, **42**, 2829-2833.

[67] Drost, F. C. and Werker, B. J. M. (1996). Closing the $GARCH$ gap : Continuous time $GARCH$ modeling. *Journal of Econometrics*, **74**, 31-57.

[68] Duflo, M. (1990). *Méthodes récursives aléatoires*. Techniques stochastiques. Massons, Paris.

[69] Durbin, J. (1959). Efficient estimation of parameters in moving average models. *Biometrika*, **46**, 306-316.

[70] Engle, R. F. (1982). Autoregressive Conditional Heteroskedasticity with estimates of variance of U.K. Inflation. *Econometrica*, **50**, 987-1008.

[71] Francq, C. and Zakoïan, J. M. (2004). Maximum likelihood estimation of pure $GARCH$ and $ARMA$-$GARCH$ processes. *Bernoulli*, **10**, 605-637.

[72] Franke, J. (1985). A Levinson-Durbin recursion for autoregressive moving average process. *Biometrika*, **72**, 573-581.

[73] Franses, P. H. (1996). *Periodicity and Stochastic Trends in Economic Times Series*. Oxford University Press.

[74] Franses, P.. and Paap, R. (2004). *Periodic time series models*. Oxford University Press.

[75] Franses, P. and Paap., R., (2000). Modeling Day-of-the-Week Seasonality in the S&P 500 Index. *Applied Financial Economics*, **10**, 483-488.

[76] Friedlander, B. (1982a). A modified prefilter for some recursive parameter estimation algorithms. *IEEE Trans. Auto. Control*, **27**, 232-235.

[77] Friedlander, B. (1982b). System identification techniques for adaptive signal processing. *IEEE Trans. Acoust. Speech Signal Process.*, **30**, 240-245.

[78] Fuller, W. A. (1976). *Introduction to Statistical Time Series*. John Wiley & Sons, New York.

[79] Gardner W. A and Franks, L. E. (1975). Characterization of cyclostationary. random signal processes. *IEEE Trans. Inform. Theory*, **21**, 4-14.

[80] Gardner, G., Harvey, A. C. and Phillips, G. D. A. (1980). Algorithm AS 154. An algorithm for exact maximum likelihood estimation of autoregressive-moving average models by means of Kalman filtering. *Applied Statistics*, **29**, 311-322.

[81] Ghysels, E. (1994). On the periodic structure of business cycle. *Journal of Business and Economic Statistics*, **12**, 289-298.

[82] Ghysels, E. (1997). On seasonality and business cycle duration : A nonparametric investication. *Journal of Econometrics*, **79**, 269-290.

[83] Giraitis, L., Kokoskza, P., and Leipus, R. (2000). Stationary $ARCH$ models: Dependence structure and central limit theorem. *Econometric Theory*, **16**, 3-22.

[84] Gladyshev, E. G. (1961). Periodically correlated random sequences. Soviet. Math., **2**, 385-88.

[85] Goodwin, G. C. and Sin, K. S. (1984). Adaptive Filtering : Prediction and Control. Englewood Cliffs, NJ: Prentice Hall.

[86] Gourieroux, C., Monfort, A. and Renault, E. (1993). Indirect Inference. *Journal of Applied Econometrics*, **3**, S85-S118.

[87] Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and its Applications*. Academic Press, New York.

[88] Hallin, M. (1984). Spectral factorization of nonstationary moving average processes. *The Annals of Statistics*, **12**, 172-192.

[89] Hallin, M. (1986). Nonstastionary q-dependent processes and time-varying moving average models : invertibility properties and the forecasting problem. *Adv. Appl. Probab.*, **18**, 170-210.

[90] Hammarling, S. J. (1982). Numerical solution of the stable, non-negative definite Lyapunov equation. IMA, *Journal of Numerical Analysis*, **2**, 303–323.

[91] Hannan, E. J. (1970). *Multiple time series*. John Wiley, New York.

[92] Hannan, E. J. (1976). The convergence of some recursions. T*he Annals of Statistics*, **4**, 1258-1270.

[93] Hannan, E. J. (1980). Recursive estimation based on $ARMA$ models. *The Annals of Statistics*, **8**, 762-777.

[94] Harvey, A. C. and Phillips, G. D. A. (1979). Maximum likelihood estimation of regression models with autoregressive-moving average disturbances. *Biometrika*, **66**, 49-58.

[95] Houacine, A. (1991). Regularized fast recursive least squares algorithms for adaptive filtering. *IEEE Trans. Signal. process.*, **39**, 860-870.

[96] Houacine, A. and Demoment, G. (1986). Chandrasekhar adaptive regularizer for adaptive filtering. *Proc. IEEE ICASSP*, Tokyo, Japan.

[97] Hui, Y. V. and Li, W. K. (1995). On fractionally differenced periodic processes. *Sankhya*, **57**, 19-31.

[98] Hurd, H. L., Makagon, A. and Miamee, A.G. (2002). On $AR(1)$ models with periodic and almost periodic coefficients. *Stochastic Processes and their Applications*, **100**, 167-185.

[99] Jones, R. H. and Brelsford, W. M. (1967). Time series with periodic structure. *Biometrika*, **54**, 403-408.

[100] Jymenez, C., McLeod, A. I. and Hipel, K. W. (1989). Kalman filter estimation for periodic autoregressive-moving average models. *Stochastic Hydrology and Hydraulics*, **3**, 227-240.

[101] Kalman, E. (1960). A new approach to linear filtering and predicting problems. *Transactions of the ASME, Journal of Basic Engineering Series*, **82**, 35-45.

[102] Kailath, T. (1968). An innovations approach to least squares estimation- Part I : Linear filtering in additive white noise, *IEEE, Trans. Autom. Contr.*, **13**, 646-654.

[103] Kesten, H., and Spitzer, F. (1984). Convergence in distribution for products of random matrices. *Z. Wahrsch. Verw. Gebiete*, **67**, 363-386.

[104] Kitagawa, G. (1977). An algorithm for solving the matrix equation $X = FXF' + S$. *Int. J. Control*, **25**, 745–753.

[105] Klein, A. and Mélard, G. (1989). On algorithms for computing the covariance matrix of estimates in autoregressive-moving average models. *Computational Statistics Quarterly*, **5**, 1-9.

[106] Klein, A. and Mélard, G. (1990). Fisher's information matrix for seasonal autoregressive-moving average models. *Journal of Time Series Analysis*, **11**, 231-237.

[107] Klein, A., Mélard, G. and Zahaf, T. (1998). Computation of the exact information matrix of Gaussian dynamic regression time series models. *The Annals of Statistics*, **26**, 1636-1650.

[108] Klein, A. and Mélard, G. (2004). An algorithm for computing the asymptotic Fisher information matrix for seasonal SISO models. *Journal of Time Series Analysis*, **25**, 627-648.

[109] Kohn, R. and Ansley, C. F. (1985). Computing the likelihood and its derivatives for a Gaussian $ARMA$ models. *Journal of Statistical Computation and Simulation*, **22**, 229-263.

[110] Kubrusly, C. S. (1985). Mean square stability for dicrete bounded linear systems in Hilbert space. *SIAM J. Control and Optimization*, **23**, 19-29.

[111] Kushner, H. J. and Clark, D. S. (1978). *Stochastic approximation methods for constrained and unconstrained systems, Applied Mathematics Science*. Springer Verlag, Berlin.

[112] Kushner, H. J. and Yin, G. G. (1997). *Stochastic approximation algorithms and applications.* Springer, New York.

[113] Lewis, P. W. and Bonnie, K. R. (2002). Nonlinear modelling of periodic threshold autoregressions using Tsmars. *Journal of Time Series Analysis*, **23**, 459-471.

[114] Li, W. K and Hui, Y. V. (1988). An Algorithm for the exact likelihood of periodic autoregressive moving average models. *Communication in Statistics- Simulation and computation*, **16**, 1483-1494.

[115] Li, W. K., Ling, S. and McAleer, M. (2002). Recent theoretical results for time series models with $GARCH$ errors. *Journal of Economic Surveys*, **16**, 245-269.

[116] Li, W. K., Ling, S. and Wong, H. (2001). Estimation for partially nonstationary multivariate autoregressive models with conditional heteroscedasticity. *Biometrika*, **8**, 1135-1152.

[117] Ling, S. (1999*a*). On probability properties of a double threshold $ARMA$ conditional heteroskedasticity model. *Journal of Applied Probability*, **36**, 688-705.

[118] Ling, S. (1999*b*). On the stationarity and the existence of moments of conditional heteroskedastic $ARMA$ models. *Statistica Sinica*, **9**, 1119-1130.

[119] Ling, S. and Deng, W. C. (1993). Parametric estimate of multivariate autoegressive models with conditional heterocovariance matrix errors. *Acta Math. Appl. Sinica*, **16**, 517-533.

[120] Ling, S., Li, W. K. (1997). Fractional autoregressive integrated moving-average time series conditional heteroskedasticity. *Journal of the American Statistical Association*, **92**, 1184-1194.

[121] Ling, S., McAleer, M. (2002). Necessary and sufficient moment conditions for the $GARCH(p,q)$ and asymmetric power $GARCH(p,q)$ models. *Econometric Theory*, **18**, 722-729.

[122] Ling, S. and McAleer, M. (2003). Asymptotic theory for a new vector *ARMA-GARCH* model. *Econometric Theory*, **19**, 280-310.

[123] Ljung, L. (1977*a*). On positive real transfer functions and the convergence of some recursive schemes. *IEEE Trans. Autom. Control.* **22** 539-551.

[124] Ljung, L. (1977*b*). Analysis of recursive stochastic algorithms. *IEEE Trans. Autom. Control.* **22**, 551-575.

[125] Ljung, L. (1999). *System identification: theory for the user.* 2nd ed., PTR Prentice Hall, Upper Saddle River, N.J.

[126] Ljung, L. and Söderström, T. (1983). *Theory and practice of recursive identification.* MIT Press, Cambridge, Massachusetts.

[127] Ljung, G. M. and Box, G. E. P. (1979). The likelihood function of stationary autoregressive-moving average models. *Applied. Statistics*, **24**, 255-259

[128] Lukacs, E. (1975). *Stochastic convergence.* 2nd ed. Academic.

[129] Lund, R. and Basawa, I. V. (1999). Modeling and inference for periodically correlated time series. *In : Gosh, S. (Ed.), Asymptotics, Nonparametrics and Time series*, Marcel Dekker, New York, 37-62.

[130] Lund, R., and Basawa, I. V. (2000). Recursive prediction and likelihood evaluation for periodic *ARMA* models. *Journal of time series analysis*, **21**, 75-93.

[131] Lütkepohl, H. (1991). *Introduction to multiple time series analysis.* Springer Verlag, Berlin.

[132] Magnus, J. R. and Neudecker H. (1989). *Matrix Differential Calculus.* John Wiley, New York.

[133] Mauricio, J. A. (2002). An algorithm for the exact likelihood of a stationary vector autoregressive-moving average model. *Journal of Time Series Analysis*, **23**, 473-486.

[134] McLeod, A. I. (1975). The Derivation of the theoretical autocovariances function of autoregressive-moving average time series. *Applied Statistics*, **24**, 255-256.

[135] McLeod, A. I. (1984). Duality and other properties of multiplicative seasonal autoregressive moving average models. *Biometrika*, **71**, 207–211.

[136] McLeod, A. I. (1993). Parsimony, model adequacy and periodic correlation in forecasting time series. *International Statistical Review*, **61**, 387-393.

[137] McLeod, A. I. (1994). Diagnostic Checking Periodic Autoregression Models with Application. *Journal of Time Series Analysis*, **15**, 221-233.

[138] McLeod, A. I. (1999). Necessary and sufficient condition for nonsingular Fisher information matrix in *ARMA* and fractional *ARMA* models. *The American Statistician*, **53**, 71-72.

[139] Mélard, G. (1984). Algorithm AS197: A fast algorithm for the exact likelihood of autoregressive-moving average models. *Applied Statistics*, **33**, 104-114.

[140] Mélard, G. (1985). *Analyse de Données chronologiques*. Séminaire de Mathémathiques supérieurs, Séminaire scientifique OTAN (NATO Advanaced Study Institute), Presses de l'université de Montréal, Montréal.

[141] Mélard, G., Klein, A. (1994). On a fast algorithm for the exact information matrix of a Gaussian ARMA time series. *IEEE Trans. Signal Process.*, **42**, 2201-2203.

[142] Meyer, R. A. and Burrus, C. S. (1975). A unified analysis of multirate and periodically timevarying digital filters. *IEEE Trans. Circuits and Systems*, **22**, 162-168.

[143] Milhoj, A. (1985). The moment structure of *ARCH* processes. *Scandinavian Journal of Statistics*, **12**, 281-292.

[144] Moore, J. B. and Ledwich, G. (1980). Multivariable adaptive parameter and state estimators with convergence analysis. *Journal of the Australian Mathematical Society*, **21**, 176-197.

[145] Moore, J. B. and Weiss, H. (1979). Recursive prediction error methods for adaptive estimation. *IEEE Trans. Systems. Man, Cyber.,* **9**, 197-205.

[146] Morf, M., Sidhu, G. S. and Kailath, T. (1974). Some new algorithms for recursive estimation on constant, linear, discrete-time systems, *IEEE Trans. Autom. Control*, **19**, 315-323.

[147] Moulines, E., Priouret, P. and Roueff, F. (2005). On recursive estimation for locally stationary time varying autoregressive processes. *The Annals of Statistics*, **33**, No. 6.

[148] Nelson, D. B. (1990). *ARCH* Models as Diffusion Approximations. *Journal of Econometrics*, **45**, 7-38.

[149] Nelson, D. B., and Foster, D. P. (1994). Asymptotic filtering theory for univariate *ARCH* models. *Econometrica*, **62**, 1-41.

[150] Neveu, J. (1975). *Discrete Parameter Martingales*. New York, North-Holland.

[151] Newton, H. J. (1978). The information matrices of the parameters of multiple mixed time series. *Journal of Multivariate Analysis*, **8**, 317-323.

[152] Newton, H. J. (1982). Using periodic autoregressions for multiple spectral estimation. *Technometrics*, **6**, 1310-1317.

[153] Nicholls, D. F. and Quinn, B. G. (1982). *Random coeffcient autoregressive model : An introduction*. Springer Verlag, New York.

[154] Osborn, D. R. (1992). The implication of periodically varying coefficients for seasonal time-series processes. *Journal of Econometrics*, **48**, 373-384.

[155] Pagano, M. (1978). On periodic and multiple autoregression. *The Annals of Statistics*, **6**, 1310-1317.

[156] Panuska, V. (1968). A stochastic approximation method for identification of linear systems using adaptive filtering. *In Joint Automatic Control Conference*, 1014-1021. Ann. Arbor. Mich.

[157] Pantula, S. G. (1988). Estimation of autoregressive models with *ARCH* errors. *Sankhya*, **50**, 119-138.

[158] Pearlman, J. G. (1980). An algorithm for the exact likelihood of a high-order autoregressive-moving average process, *Biometrika*, **67**, 232-233.

[159] 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.

[160] Plackett, R. L. (1950). Some theorems in least squares. *Biometrika*, **32**, 149.

[161] Pollock, D. S. G. (2003). Recursive estimation in econometrics. *Computational Statistics and Data Analysis*, **44**, 37-75.

[162] Porat, B. and Friedlander, B. (1986*a*). Bounds on accuracy of Gaussian *ARMA* parameter estimation methods based on sample covariance, *IEEE Trans. Auto. Control*, **31**, 579-582.

[163] Porat, B. and Friedlander, B. (1986*b*). Computation of the exact information matrix of Gaussian time series with stationary random components, *IEEE Trans. Acoust. Speech. Signal Process*, **14**, 118-130.

[164] Poskitt, D. S. and Treymane, A. R. (1981). An approach to testing linear time series models. *The Annals of Statistics*, **9**, 974-986.

[165] Risanen, J. (1973). Algorithms for trianguar decomposition of block Hankel and Toeplitz matrices with application to factoring positive matrix polynomials. *Mathematics of computation*, **27**, 147-154.

[166] Robbins, H. and Monro, S. (1951). A stochastic approximation method. *Annals of Mathematical Statistics*, **22**, 702-710.

[167] Robbins, H., and Siegmund, D. (1971). A convergence theorem for nonnegative almost supermartingales and some applications. *In Optimizing Meth. Statist., Proc. Sympos (p.233-257).* Ohio State University.

[168] Sakai, H. (1982). Circular Lattice Filtering Using Pagano's Method. IEEE Trans. *Acoustics, Speech, and Signal Processing*, **30**, 279-286.

[169] Salas, J. D, Duane C. Boes and Smith, R. A. (1983). Estimation of *ARMA* Models with Seasonal Parameters. *Water Resources Research*, **18**, 1006-1010.

[170] Sen P. K.and J. M. Singer, (1993). *Large sample methods in statistics* : *an introduction with applications*. Chapman & Hall, London.

[171] Shao, Q. (2005). On mixture periodic autoregresive models. Preprint.

[172] Shao, Q. and Lund, R. (2004). Computation and characterization of autocorrelations and partial autocorrelations in periodic ARMA models. *Journal of Time Series Analysis*, **25**, 359-372.

[173] Shea, B. L. (1989). The exact likelihood of a vector autoregressive moving average model. *Journal of the Royal Statistical Society* C, **38**, 161-184.

[174] Söderström, 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.

[175] Solo, V. (1978). "Time series recursions and stochastic approximation". Ph.D. Thesis submitted to Australian National Univ.

[176] Solo, V. (1979). The convergence of AML. *IEEE Trans. Autom. Control*. **24**, 958-963.

[177] Solo, V. (1980). Some aspects of recursive parameter estimation. *International Journal of control*, **32**, 395-410.

[178] Solo, V. (1981). The second order properties of a time series recursion. *The Annals of Statistics*, **9**, 307-317.

[179] Solo, V. and Kong, X. (1995). *Adaptive Signal Processing Algorithms*. Prentice-Hall, Englewood Cloes, NJ.

[180] Sreedhar, J. and Van Dooren, P. (1994). A Schur approach for solving some periodic matrix equations. *In: Systems and Networks: Mathematical Theory and Applications (U. Helmke,R. Mennicken and J. Saurer, Eds.)*. Vol. 77 of Mathematical Research. 339-362.

[181] Stout, W. (1974). *Almost Sure Convergence*. New York: Academic Press.

[182] Tiao, G. C. and Grupe, M. R. (1980). Hidden periodic autoregressive-moving average models in time series data. *Biometrika*, **67**, 365-373.

[183] Tol, R. J. (1996). Autoregressive conditional heteroscedasticity in daily temperature measurments, *Environmetrics*, **7**, 67-76.

[184] Tong, H. (1990). *Nonlinear Time Series : a Dynamical System Approach*. Oxford University Press, Oxford.

[185] Troutman, B. M. (1979). Some results in periodic autoregression. *Biometrika*, **66**, 219-227.

[186] Tsay R. S. (1987). Conditional heteroskedastic time series models. *Journal of the American Statistical Association*, **7**, 590-604.

[187] Tweedie, R. L. (1988). Invariant measure for Markov chains with no irreducibility assumptions. *Journal of Applied Probability*, **25**, 275-285.

[188] Ula, T. A. (1990). Periodic covariance stationarity of multivariate periodic autoregressive moving average processes. *Water Resour. Res.*, **26**, 855-861.

[189] Ula, T. A. and Smadi, A. A. (1997). Periodic stationary conditions for periodic autoregressive moving average processes as eigenvalue problems. *Water Resources Research*, **33**, 1929-1934.

[190] Ula, T. A. and Smadi, A. A. (2003). Identication of periodic moving average models. *Communications in Statistics- Theory and Methods*, **32**, 2465-2475.

[191] Varga, A. (1997). Periodic Lyapunov equations: some applications and new algorithms. *Int. J. Control*, **67**, 69-87.

[192] 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.

[193] Vecchia, A. V. (1985a). Maximum likelihood estimation for periodic autoregressive Moving average models. *Technometrics*, **27**, 375-384.

[194] Vecchia, A. V. (1985b). Periodic autoregressive-moving average ($PARMA$) modeling with application to water resources. *Water Resources Bulletin*, **21**, 721-730.

[195] Vecchia, A. V. and Ballerini, R. (1991). Testing for periodic autocorrelations in seasonal time series data. *Biometrika*, **78**, 53-63.

[196] Vecchia, A. V., Obeysekera, J. T. and Salas, J. D. (1983). Aggregation and estimation for low-order periodic ARMA models. *Water Resources Research*, **19**, 1297-1306.

[197] Wang, D., Song, L. and Shi, S. (2004). Estimation and testing for the parameters of $ARCH(q)$ under ordered restriction. *Journal of Time Series Analysis*, **25**, 483-499.

[198] Weiss, A. A. (1986). Asymptotic theory for $ARCH$ models : estimation and testing. *Econometric Theory*, **2**, 107-131.

[199] Whittle, P. (1953). Estimation and information in time series. *Ark. Math. Fys. Astr.*, **2**, 433-434.

[200] Wilson, G. T. (1979). Some efficient computational procedures for high order *ARMA* models. *Journal of Statistical Computation and Simulation*, **8**, 303-309.

[201] Wold, H. (1938). *A study in the stationary time series.* Almqvist and Wiksell, Stockholm (2nd edition : 1954).

[202] Wong, H. and Li, W. K., (1997). On multivariate conditional heteroskedasticity model. *Biometrika*, **4**, 111-123.

[203] Yang Z. P. and Feng, X. X. (2002). A note on the trace inequality for products of hermitian matrix power. *Journal of Inequalities in Pure and Appllied Mathematics*, **3**, A. 78.

[204] Young, P. C. (1974). Recursive approaches to time series analysis. *Bull. Inst. Math. Appl.*, **10**, 209-224.

[205] Young, P. C. (1984). *Recursive Estimation and Time-Series Analysis, An Introduction.* Springer-Verlag.

[206] Zahaf, T. (1998). "Contribution à l'estimation des paramètres de séries chronologiques". Thèse de Doctorat es siences. Institut de Statistique et de Recherche Opérationnelle, Université Libre de Bruxelles.

Résumé

L'objectif de cette thèse est l'étude de la modélisation des séries chronologiques exhibant une dynamique périodique. Nous considérons deux types de formulations dont l'une est linéaire, s'exprimant à travers des équations aux différences stochastiques à coefficients périodiques (scalaires et matricielles) et l'autre non linéaire du type conditionnellement hétéroscédastique. Notre contribution est triple :

D'abord, l'étude de certaines propriétés algébriques et probabilistes de ces modèles, à savoir : la stabilité (causalité et inversibilité), la stationnarité périodique stricte, la structure d'autocorrélation, l'information de Fisher et l'existence des moments d'ordres supérieurs.

Ensuite, l'élaboration des méthodes d'estimation basées sur le critère du maximum de vraisemblance et celui des moindres carrés (ordinaire, pondéré et en deux étapes). Certaines méthodes sont restreintes au cas univarié et d'autres s’appliquent également dans le cas multivarié. De plus l'estimation est abordée selon deux contextes différents, à savoir l'estimation hors-ligne et l’estimation en-ligne.

Et enfin, l'étude des propriétés statistiques (asymptotiques) des estimateurs fournis par les méthodes développées, telles que la convergence presque sure, la normalité asymptotique et l'efficacité asymptotique.